LECTURE NOTES IN COMPUTATIONAL SCIENCE AND ENGINEERING 91

Randolph Bank · Michael Holst
Olof Widlund · Jinchao Xu *Editors*

Domain Decomposition Methods in Science and Engineering XX

Springer

Lecture Notes
in Computational Science
and Engineering

91

Editors:

For further volumes:
http://www.springer.com/series/3527

Randolph Bank • Michael Holst • Olof Widlund
Jinchao Xu
Editors

Domain Decomposition Methods in Science and Engineering XX

Springer

Editors
Randolph Bank
Michael Holst
Department of Mathematics
University of California
San Diego
La Jolla
CA, USA

Jinchao Xu
Department of Mathematics
Pennsylvania State University
State College
PA, USA

Olof Widlund
Courant Institute
New York University
NY, USA

ISSN 1439-7358
ISBN 978-3-642-35274-4 ISBN 978-3-642-35275-1 (eBook)
DOI 10.1007/978-3-642-35275-1
Springer Heidelberg New York Dordrecht London

Library of Congress Control Number: 2013942186

Math. Subj. Class. (2010): 65F10, 65N30, 65N55

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

Domain decomposition, a form of divide and conquer for mathematical problems posed over a physical domain, as in partial differential equations, is the most common paradigm for large-scale simulation on massively parallel distributed, hierarchical memory computers. In domain decomposition, a large problem is reduced to a collection of (typically many) smaller problems, each of which is easier to solve computationally than the undecomposed problem and most or all of which can be solved independently and concurrently. Typically, it is necessary to iterate over the collection of smaller problems, and much of the theoretical interest in domain decomposition algorithms lies in ensuring that the number of iterations required is very small. Indeed, the best domain decomposition methods share with their cousins, multigrid methods, the property that the total computational work is linearly proportional to the size of the input data or that the number of iterations required is at most logarithmic in the number of degrees of freedom of individual subdomains. Algorithms whose work requirements are linear in the size of the input data in this context are said to be "optimal." Optimal domain decomposition algorithms are now known for many, but certainly not all, important classes of problems that arise from science and engineering. Much of the practical interest in domain decomposition algorithms lies in extending the classes of problems for which optimal algorithms are known. Domain decomposition algorithms can be tailored to the properties of the physical system as reflected in the mathematical operators, the number of processors available, and even to specific architectural parameters, such as cache size and the ratio of memory bandwidth to floating-point processing rate.

Since the first meeting was held in Paris in 1987, the International Conference on Domain Decomposition Methods is the only regularly occurring international forum dedicated to interdisciplinary technical interactions between theoreticians and practitioners working in the creation, analysis, software implementation, and application of domain decomposition methods. The conferences have now been held in 12 countries in the Far East, Europe, the Middle East, and North America. To date, there are essentially no real alternatives to domain decomposition as a strategy for parallelization on petascale computers and beyond, with hundreds of thousands or even millions of processor cores. Domain decomposition has proved to be an ideal

paradigm not only for execution on advanced architecture computers but also for the development of reusable, portable software. The most complex operation in a typical domain decomposition method is the application of a preconditioner that carries out in each subdomain step nearly identical to those required to apply a conventional preconditioner to the global domain. Hence, software developed for the global problem can readily be adapted to the local problem, instantly presenting wealth of "legacy" scientific code to be harvested for parallel implementations. Furthermore, since the majority of data sharing between subdomains in domain decomposition codes occurs in two archetypal communication operations – ghost point updates in overlapping zones between neighboring subdomains and global reduction operations, as in forming an inner product – domain decomposition methods map readily onto optimized, standardized message-passing environments, such as MPI. Finally, it should be noted that domain decomposition is often a natural paradigm for the modeling community. Physical systems are often decomposed into two or more contiguous subdomains based on phenomenological considerations, such as the importance or negligibility of viscosity or reactivity, or any other feature, and the subdomains are discretized accordingly, as independent tasks. This physically based domain decomposition may be mirrored in the software engineering of the corresponding code, and leads to threads of execution that operate on contiguous subdomain blocks, which can either be further subdivided or aggregated to the granularity of an available parallel computer, and have the correct topological and mathematical characteristics for scalability. Much of the reputation of this conference series results from the close interaction between experts in mathematics, computer science, and large-scale computational science in various application areas.

This volume contains a selection of 83 papers presented at the 20th International Conference on Domain Decomposition, DD20, hosted by the Center for Computational Mathematics at the University of California at San Diego, held at the San Diego Supercomputer Center on the UCSD campus during the week of February 9–13, 2011. The conference featured 16 plenary lectures delivered by leaders in the field, 18 minisymposiums, as well as contributed talks and a poster session. In addition, Olof Widlund gave an introductory short course on domain decomposition on Sunday February 8 to a packed room of more than 40 participants in the Center for Computational Mathematics, a short walk from the San Diego Supercomputer Center. Attending the regular conference during the week were 199 scientists from 21 countries, giving a total of 173 presentations, which accentuates the international scope and relevance of this meeting. To add a unique local flavor to the UCSD meeting, three special plenary talks were scheduled for Tuesday, given by world-renowned local UCSD computational scientists in fields spanning computational chemistry to galaxy collision simulation. In addition to the scientific talks during the day throughout the week, participants gathered for a poster session with wine and cheese in the early evening on Monday, and the plenary speakers gathered for a small dinner in Del Mar on Tuesday evening. The Scientific Committee met with the local organizing committee and discussed plans for the next conference in the series on Wednesday evening, aided by samplings from local San Diego microbreweries.

The large conference banquet for all the participants was held in the UCSD Faculty Club on Thursday evening, and the conference came to a close at noon on Friday.

For further information, we recommend the homepage of International Domain Decomposition Conferences, www.ddm.org, maintained by Martin Gander. This site features free online access to the proceedings of all previous DD conferences, information about past and future meetings, as well as bibliographic and personal information pertaining to domain decomposition. A bibliography with all previous proceedings is provided below, along with some major review articles and monographs. (We apologize for unintentional omissions to our necessarily incomplete list.) No attempts have been made to supplement this list with the larger and closely related literature of multigrid and general iterative methods, except for the books by Hackbusch and Saad, which have significant domain decomposition components.

The editors wish to thank all members of the International Scientific Committee for Domain Decomposition Conferences, chaired by Ralf Kornhuber, for their help in setting the scientific direction of this conference. We are also grateful to the organizers of the minisymposiums for shaping the profile of the scientific program and attracting high-quality presentations. The local organizers were Randolph Bank and Michael Holst, aided by Rob Falgout, David Keyes, Rich Lehoucq, and Jinchao Xu. We gratefully acknowledge administrative assistance from the San Diego Computer Center (SDSC) and the California Institute for Telecommunications and Information Technology (CalIT2).

DD20 was financially supported by the National Science Foundation, the US Department of Energy, Lawrence Livermore and Sandia National Laboratories, SDSC, CalIT2, the National Biomedical Computation Resource, and the University of California at San Diego. Finally, we would like to thank Martin Peters and Thanh-Ha Le Thi of Springer for their friendly and efficient collaboration in the production of this proceedings volume.

Randolph E. Bank
University of California, San Diego, USA

Michael J. Holst
University of California, San Diego, USA

Olof B. Widlund
Courant Institute, New York, USA

Jinchao Xu
Pennsylvania State University, USA

Proceedings from Prior Conferences in the DD Series

[DD01] Roland Glowinski, Gene H. Golub, Gérard A. Meurant, and Jacques Périaux, editors. *First International Symposium on Domain Decomposition Methods for Partial Differential Equations*. Society for Industrial and Applied

Mathematics (SIAM), Philadelphia, PA, 1988. Proceedings of the symposium held in Paris, France, January 7–9, 1987.

[DD02] Tony F. Chan, Roland Glowinski, Jacques Périaux, and Olof B. Widlund, editors. *Domain decomposition methods*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1989. Proceedings of the Second International Symposium held at the University of California, Los Angeles, California, January 14–16, 1988.

[DD03] Tony F. Chan, Roland Glowinski, Jacques Périaux, and Olof B. Widlund, editors. *Third International Symposium on Domain Decomposition Methods for Partial Differential Equations*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1990. Proceedings of the symposium held in Houston, Texas, March 20–22, 1989.

[DD04] Roland Glowinski, Yuri A. Kuznetsov, Gérard Meurant, Jacques Périaux, and Olof B. Widlund, editors. *Fourth International Symposium on Domain Decomposition Methods for Partial Differential Equations*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1991. Proceedings of the symposium held in Moscow, USSR, May 21–25, 1990.

[DD05] David E. Keyes, Tony F. Chan, Gérard Meurant, Jeffrey S. Scroggs, and Robert G. Voigt, editors. *Fifth International Symposium on Domain Decomposition Methods for Partial Differential Equations*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1992. Proceedings of the symposium held in Norfolk, Virginia, May 6–8, 1991.

[DD06] Alfio Quarteroni, Jacques Périaux, Yuri A. Kuznetsov, and Olof B. Widlund, editors. *Domain decomposition methods in science and engineering*, volume 157 of *Contemporary Mathematics*. American Mathematical Society, Providence, RI, 1994. Proceedings of the Sixth International Conference on Domain Decomposition held in Como, Italy, June 15–19, 1992.

[DD07] David E. Keyes and Jinchao Xu, editors. *Domain decomposition methods in scientific and engineering computing*, volume 180 of *Contemporary Mathematics*. American Mathematical Society, Providence, RI, 1994. Proceedings of the Seventh International Conference on Domain Decomposition held at the Pennsylvania State University, University Park, Pennsylvania, October 27–30, 1993.

[DD08] Roland Glowinski, Jacques Périaux, Zhong-Ci Shi, and Olof Widlund, editors. *Domain decomposition methods in sciences and engineering*. John Wiley & Sons Ltd., Chichester, 1997. Proceedings of the 8th International Conference held in Beijing, China, May 16–20, 1995.

[DD09] Petter E. Bjørstad, Magne S. Espedal, and David E. Keyes, editors. *Proceedings of the 9th International Conference on Domain Decomposition Methods in Bergen, Norway*. DDM.org, Augsburg, 1996. Held in Bergen, Norway, June 4–7, 1996.

[DD10] Jan Mandel, Charbel Farhat, and Xiao-Chuan Cai, editors. *Domain decomposition methods 10*, volume 218 of *Contemporary Mathematics*. American Mathematical Society, Providence, RI, 1998. Proceedings of the Tenth Interna-

tional Conference on Domain Decomposition Methods held at the University of Colorado, Boulder, Colorado, August 10–14, 1997.

[DD11] Choi-Hong Lai, Petter E. Bjørstad, Mark Cross, and Olof Widlund, editors. *Eleventh International Conference on Domain Decomposition Methods*. DDM.org, Augsburg, 1999. Proceedings of the conference held at the University of Greenwich, London, UK, July 20–24, 1998.

[DD12] Tony Chan, Takashi Kako, Hideo Kawarada, and Olivier Pironneau, editors. *Domain decomposition methods in sciences and engineering*. DDM.org, Augsburg, 2001. Proceedings of the 12th International Conference on Domain Decomposition Methods held at Chiba University, Chiba, Japan, October 25–29, 1999.

[DD13] Naima Debit, Marc Garbey, Ronald Hoppe, David Keyes, Yuri Kuznetsov, and Jacques Périaux, editors. *Domain decomposition methods in science and engineering*. Theory and Engineering Applications of Computational Methods. International Center for Numerical Methods in Engineering (CIMNE), Barcelona, 2002. Papers from the 13th International Conference on Domain Decomposition Methods held in Lyon, France, October 9–12, 2000.

[DD14] Ismael Herrera, David E. Keyes, Olof B. Widlund, and Robert Yates, editors. *Domain decomposition methods in science and engineering*. National Autonomous University of Mexico (UNAM), México, 2003. Papers from the 14th International Conference on Domain Decomposition Methods held at the Universidad Nacional Autónoma de México, Cocoyoc, January 6–12, 2002.

[DD15] Ralf Kornhuber, Ronald Hoppe, Jacques Périaux, Olivier Pironneau, Olof Widlund, and Jinchao Xu, editors. *Domain decomposition methods in science and engineering*, volume 40 of *Lecture Notes in Computational Science and Engineering*. Springer-Verlag, Berlin, 2005. Papers from the 15th International Conference on Domain Decomposition held at the Freie Universität Berlin, Berlin, Germany, July 21–25, 2003.

[DD16] Olof B. Widlund and David E. Keyes, editors. *Domain decomposition methods in science and engineering XVI*, volume 55 of *Lecture Notes in Computational Science and Engineering*. Springer, Berlin, 2007. Papers from the 16th International Conference on Domain Decomposition Methods held in New York, USA, January 11–15, 2005.

[DD17] Ulrich Langer, Marco Discacciati, David E. Keyes, Olof B. Widlund, and Walter Zulehner, editors. *Domain decomposition methods in science and engineering XVII*, volume 60 of *Lecture Notes in Computational Science and Engineering*. Springer-Verlag, Berlin, 2008. Selected papers from the 17th International Conference on Domain Decomposition Methods (DD17) held in Strobl, Austria, July 3–7, 2006.

[DD18] Michel Bercovier, Martin J. Gander, Ralf Kornhuber, and Olof Widlund, editors. *Domain decomposition methods in science and engineering XVIII*, volume 70 of *Lecture Notes in Computational Science and Engineering*. Springer-Verlag, Berlin, 2009. Selected papers from the 18th International Conference held at the Hebrew University of Jerusalem, Jerusalem, Israel, January 12–17, 2008.

[DD19] Yunqing Huang, Ralf Kornhuber, Olof Widlund, and Jinchao Xu, editors. *Domain decomposition methods in science and engineering XIX*, volume 78 of *Lecture Notes in Computational Science and Engineering*. Springer, Heidelberg, 2011. Selected papers from the 19th International Conference held in Zhanjiajie, China, August 17–21, 2009.

Additional Key DD References

[1] Tony F. Chan and Tarek P. Mathew. Domain decomposition algorithms. In *Acta numerica, 1994*, Acta Numer., pages 61–143. Cambridge Univ. Press, Cambridge, 1994.

[2] Charbel Farhat and François-Xavier Roux. Implicit parallel processing in structural mechanics. *Comput. Mech. Adv.*, 2(1):124, 1994.

[3] Wolfgang Hackbusch. *Iterative solution of large sparse systems of equations*, volume 95 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 1994. Translated and revised from the 1991 German original.

[4] David E. Keyes, Youcef Saad, and Donald G. Truhlar, editors. *Domain-based parallelism and problem decomposition methods in computational science and engineering*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1995.

[5] Boris N. Khoromskij and Gabriel Wittum. *Numerical solution of elliptic differential equations by reduction to the interface*, volume 36 of *Lecture Notes in Computational Science and Engineering*. Springer-Verlag, Berlin, 2004.

[6] V. G. Korneev and U. Langer. Domain decomposition and preconditioning. In Erwin Stein, René de Borst, and Thomas J. R. Hughes, editors, *Encyclopedia of Computational Mechanics*. John Wiley & Sons Ltd., Chichester, 2004.

[7] J. Kruis. *Domain Decomposition for Distributed Computing*. Dun Eaglais, Saxe Coburg, 2005.

[8] Ulrich Langer and Olaf Steinbach. Coupled finite and boundary element domain decomposition methods. In *Boundary element analysis*, volume 29 of *Lect. Notes Appl. Comput. Mech.*, pages 61–95. Springer, Berlin, 2007.

[9] Patrick Le Tallec. Domain decomposition methods in computational mechanics. *Comput. Mech. Adv.*, 1(2):121–220, 1994.

[10] V. I. Lebedev and V. I. Agoshkov. *Operatory Puankare-Steklova i ikh prilozheniya v analize (Poincaré-Steklov operators and their applications in analysis)*. Akad. Nauk SSSR Vychisl. Tsentr, Moscow, 1983 (in Russian).

[11] Tarek P. A. Mathew. *Domain decomposition methods for the numerical solution of partial differential equations*, volume 61 of *Lecture Notes in Computational Science and Engineering*. Springer-Verlag, Berlin, 2008.

[12] Sergey Nepomnyaschikh. Domain decomposition methods. In *Lectures on advanced computational methods in mechanics*, volume 1 of *Radon Ser. Comput. Appl. Math.*, pages 89–159. Walter de Gruyter, Berlin, 2007.

[13] Peter Oswald. *Multilevel finite element approximation*. Teubner Skripten zur Numerik. [Teubner Scripts on Numerical Mathematics]. B. G. Teubner, Stuttgart, 1994. Theory and applications.

[14] L. Pavarino and A. Toselli. *Recent developments in Domain Decomposition Methods*, volume 23 of *Lecture Notes in Computational Science and Engineering*. Springer-Verlag, Berlin, 2002.

[15] Alfio Quarteroni and Alberto Valli. *Domain decomposition methods for partial differential equations*. Numerical Mathematics and Scientific Computation. The Clarendon Press Oxford University Press, New York, 1999. Oxford Science Publications.

[16] Yousef Saad. *Iterative methods for sparse linear systems*. Society for Industrial and Applied Mathematics, Philadelphia, PA, second edition, 2003.

[17] Barry F. Smith, Petter E. Bjørstad, and William D. Gropp. *Domain decomposition*. Cambridge University Press, Cambridge, 1996. Parallel multilevel methods for elliptic partial differential equations.

[18] Olaf Steinbach. *Stability estimates for hybrid coupled domain decomposition methods*, volume 1809 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 2003.

[19] Andrea Toselli and Olof Widlund. *Domain decomposition methods—algorithms and theory*, volume 34 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin, 2005.

[20] Barbara I. Wohlmuth. *Discretization methods and iterative solvers based on domain decomposition*, volume 17 of *Lecture Notes in Computational Science and Engineering*. Springer-Verlag, Berlin, 2001.

[21] Jinchao Xu. Iterative methods by space decomposition and subspace correction. *SIAM Rev.*, 34(4):581–613, 1992.

[22] Jinchao Xu and Jun Zou. Some nonoverlapping domain decomposition methods. *SIAM Rev.*, 40(4):857–914, 1998.

Contents

Part I

Plenary Presentations

Equidistribution and Optimal Approximation Class*

Constantin Bacuta[1], Long Chen[2], and Jinchao Xu[3]

[1] Department of Mathematical Sciences, University of Delaware, Newark, DE 19716
bacuta@math.udel.edu
[2] Department of Mathematics, University of California at Irvine, Irvine, CA 92617
chenlong@math.uci.edu
[3] Department of Mathematics, Pennsylvania State University, University Park, PA 16802
xu@math.psu.edu

1 Introduction

Local adaptive grid refinement is an important technique in finite element methods. Its study can be traced back to the pioneering work [2] in one dimension. In recent years, mathematicians start to prove the convergence and optimal complexity of the adaptive procedure in multi-dimensions. Dörfler [11] first proved an error reduction in the energy norm for the Poisson equation provided the initial mesh is fine enough. Morin et al. [15, 16] extended the convergence result without the constrain of the initial mesh and they also reveal the importance of data oscillation. But results in [11, 15, 16] only establish the qualitative convergence estimate by a proof of an error reduction property. The number of elements generated by the adaptive algorithm is not under control. A natural theoretical question is if a standard adaptive finite element scheme would give an optimal asymptotic convergence rate in terms of the number of elements. For linear finite element approximation to second order elliptic boundary value problems in two dimensions, for example, an optimal asymptotic error estimate would be something like

$$|u - u_N|_{1,\Omega} \leq C(u)N^{-1/2}, \tag{1}$$

where u_N is a finite element approximation of the Poisson equation with homogenous Dirichlet boundary condition based on an adaptive grid with at most N elements.

An important progress has been made by Binev et al. [7] concerning the asymptotic estimate (1). In their algorithm, an additional coarsening step is required to achieve optimal complexity. However in practice the nearly optimal complexity

* The work of C. Bacuta was partially supported by NSF DMS-0713125. L. Chen was supported in part by NSF Grant DMS-0811272, DMS-1115961, and in part by DOE Grant DE-SC0006903. J. Xu was supported in part by NSF DMS-0915153, NSFC-10528102, Alexander von Humboldt Research Award for Senior US Scientists and in part by DOE Grant DE-SC0006903.

R. Bank et al. (eds.), *Domain Decomposition Methods in Science and Engineering XX*, Lecture Notes in Computational Science and Engineering 91,
DOI 10.1007/978-3-642-35275-1_1,

is obtained without the coarsening step. Such theoretical gap is filled by Stevenson [18] which shows that the practical refinement without a recurrent coarsening will also generate finite element solution with quasi-optimal computational complexity. But marking for oscillation and refinement with interior nodes assumptions are still needed. Recently, [8] presented the most standard AFEM and proved a contraction property and quasi-optimal cardinality without any additional assumptions. Their results show that if the solution $u \in \mathscr{A}_s$, where $\mathscr{A}_s$ is the approximation class space of rate s, then $|u-u_N|_{1,\Omega} \leq |u|_{\mathscr{A}_s} N^{-s}$.

Another important theoretical and practical issue is to characterize the approximation class $\mathscr{A}_{1/2}$ using the smoothness of u. A near characterization of $\mathscr{A}_{1/2}$ in terms of Besov spaces $B^k_{p,q}(\Omega)$ in two dimensions can be found in [6, 7] which shows that $u \in \mathscr{A}_{1/2}$ implies that $u \in B^2_{1,1}(\Omega)$ and $u \in B^2_{p,p}(\Omega)$ for $p > 1$ implies that $u \in \mathscr{A}_{1/2}$.

In this paper, we shall provide a sharper result: We prove that
if $u \in W^{2,L\log L}(\Omega)$, i.e.,

$$\int_\Omega |D^2 u \log |D^2 u|\,|\, dx < \infty,$$

then $u \in \mathscr{A}_{1/2}$. This is an improved result since, when $p > 1$, $B^2_{p,p}(\Omega) \subset W^{2,L\log L}(\Omega)$ from the Hölder inequality. With the regularity theory of elliptic equations, which ensures $u \in W^{2,L\log L}(\Omega)$, we are led to conclude the following practical statement: linear adaptive finite element approximation of second order elliptic equations in two dimensions will achieve optimal rate of convergence.

Our contribution in this paper is further related with recent work on equidistribution and refinement strategies as follows:

1. The role of the equidistribution. In Sect. 2 we reveal that the equidistribution principle can be severely violated but asymptoticly optimal error estimates can still be maintained. The result (Theorem 1) is firstly presented in [9] and similar idea can be also found in [8] around the same time.
2. The proof of the bound of the pollution of the local mesh refinement in the completion is of its own interest. The estimate (Theorem 2) is a much sharper constant comparing with existing results in [7]. The idea of the proof is borrowed from [1] and the result is generalized from the uniform grids in [1] to compatibly divisible unstructured grids.

The rest of the paper is organized as follows. In Sect. 2 we explain the equidistribution principle for the case when the function to be approximated belongs to $W^{2,1}(\Omega)$. The advantage of our approach is that only standard approximation for the interpolation operator are used, and approximation theory for Besov spaces is not needed. In Sect. 3, we review the newest vertex bisection refinement strategy and provide a sharp estimate for the number of triangle needed for the completion of the mesh after an arbitrary marking and bisection refinement is performed. In Sect. 4, we present a new approach for the local grid refinement based on the error estimate and the equidistribution principle.

2 Error Estimate and Equidistribution Principle

We shall consider a simple elliptic boundary value problem

$$-\Delta u = f \text{ in } \Omega, \qquad u = 0 \text{ on } \partial\Omega, \tag{2}$$

where, for simplicity, we assume Ω is a polygon and is partitioned by a shape regular conforming triangulation $\mathscr{T}_N$ with N number of triangles. Let $\mathscr{V}_N \subset H_0^1(\Omega)$ be the corresponding continuous piecewise linear finite element space associated with this triangulation $\mathscr{T}_N$.

A finite element approximation of the above problem is to find $u_N \in \mathscr{V}_N$ such that

$$a(u_N, v_N) = (f, v_N) \quad \forall v_N \in \mathscr{V}_N, \tag{3}$$

where

$$a(u,v) = \int_\Omega \nabla u \cdot \nabla v dx, \text{ and } (f,v) = \int_\Omega f v dx.$$

For this problem, it is well known that for a fixed finite element space $\mathscr{V}_N$

$$|u - u_N|_{1,\Omega} = \inf_{v_N \in \mathscr{V}_N} |u - v_N|_{1,\Omega}. \tag{4}$$

We then present an H^1 error estimate for linear triangular element interpolation in two dimensions. We note that in two dimensions, the following two embeddings are both valid:

$$W^{2,1}(\Omega) \subset W^{1,2}(\Omega) \equiv H^1(\Omega) \text{ and } W^{2,1}(\Omega) \subset C(\bar{\Omega}). \tag{5}$$

Given $u \in W^{2,1}(\Omega)$, let u_I be the linear nodal value interpolant of u on $\mathscr{T}_N$. For any triangle $\tau \in \mathscr{T}_N$, thanks to (5) and the assumption that τ is shape-regular, we have

$$|u - u_I|_{1,\tau} \lesssim |u|_{2,1,\tau}.$$

As a result,

$$|u - u_I|_{1,\Omega}^2 \lesssim \sum_{\tau \in \mathscr{T}_N} |u|_{2,1,\tau}^2.$$

To minimize the error, we can try to minimize the right hand side. By Cauchy-Schwarz inequality,

$$|u|_{2,1,\Omega} = \sum_{\tau \in \mathscr{T}_N} |u|_{2,1,\tau} \le (\sum_{\tau \in \mathscr{T}_N} 1)^{1/2} (\sum_{\tau \in \mathscr{T}_N} |u|_{2,1,\tau}^2)^{1/2} = N^{1/2} (\sum_{\tau \in \mathscr{T}_N} |u|_{2,1,\tau}^2)^{1/2}.$$

Thus, we have the following lower bound:

$$(\sum_{\tau \in \mathscr{T}_N} |u|_{2,1,\tau}^2)^{1/2} \ge N^{-1/2} |u|_{2,1,\Omega}. \tag{6}$$

The equality holds if and only if

$$|u|_{2,1,\tau} = \frac{1}{N}|u|_{2,1,\Omega}. \tag{7}$$

The condition (7) is hard to be satisfied in general. But we can considerably relax this condition to ensure the lower bound estimate (6) is still achieved asymptotically. The relaxed condition is as follows:

$$|u|_{2,1,\tau} \leq \kappa_{\tau,N}|u|_{2,1,\Omega} \tag{8}$$

and

$$\sum_{\tau\in\mathscr{T}_N} \kappa_{\tau,N}^2 \leq c_1 N^{-1}. \tag{9}$$

When the above two inequalities hold, we have

$$|u-u_I|_{1,\Omega} \lesssim N^{-1/2}|u|_{2,1,\Omega}.$$

In summary, we have the following theorem.

Theorem 1. *If $\mathscr{T}_N$ is a triangulation with at most N triangles and satisfying (8) and (9), then*

$$|u-u_N|_1 \leq |u-u_I|_{1,\Omega} \lesssim N^{-1/2}|u|_{2,1,\Omega}. \tag{10}$$

In the above analysis, we see how equidistribution principle plays an important role in achieving asymptotically optimal accuracy for adaptive grids. We would like to further elaborate that, in the current setting, equidistribution is indeed a sufficient condition for optimal error, but by no means this has to be a necessary condition. Namely the equidistribution principle can be severely violated but asymptoticly optimal error estimates can still be maintained. For example, the following mild violation of this principle is certainly acceptable:

$$|u|_{2,1,\tau} \leq \frac{c}{N}|u|_{2,1,\Omega}. \tag{11}$$

In fact, this condition can be more significantly violated on a finitely many elements $\{\tau\}$

$$|u|_{2,1,\tau} \leq \frac{c}{\sqrt{N}}|u|_{2,1,\Omega}. \tag{12}$$

It is easy to see if a bounded number of elements satisfy (12) and the rest satisfy (11), the estimate (9) is satisfied and hence the optimal error estimate (10) is still valid.

As we can see that the condition (12) is a very serious violation of equidistribution principle, nevertheless, as long as such violations do not occur on too many elements, asymptotically optimal error estimates are still valid. This simple observation is important from both theoretical and practical points of view. The marking strategy proposed by Dörfler [11] may also be interpreted in this way in its relationship with equidistribution principle. In [5], Binev and DeVore propose to use certain penalty in using equidistribution principle. Such a modification certainly has similar spirit.

We shall discuss how to generate a mesh $\mathscr{T}_N$ to satisfy (8) and (9) in the next two sections. To this end, we shall introduce the local refinement method: newest vertex bisection, in the next section.

3 Newest Vertex Bisection

In this section we shall give a brief introduction of the newest vertex bisection and mainly concern the number of elements added by the completion process. We refer to [14, 19] and [7] for detailed description of the newest vertex bisection refinement procedure.

Given an initial shape regular triangulation $\mathscr{T}_0$ of Ω, it is possible to assign to each $\tau \in \mathscr{T}_0$ exactly one vertex called *the newest vertex*. The opposite edge of the newest vertex is called *refinement edge*. The rule of the newest vertex bisection includes:

1. A triangle is divided to two new children triangles by connecting the newest vertex to the midpoint of the refinement edge;
2. The new vertex created at a midpoint of a refinement edge is assigned to be the newest vertex of the children.

It is easy to verify that all the descendants of an original triangle fall into four similarity classes (see Fig. 1) and hence the angles are bounded away from 0 and π and all triangulations refined from $\mathscr{T}_0$ using newest vertex bisection forms a shape regular class of triangulations.

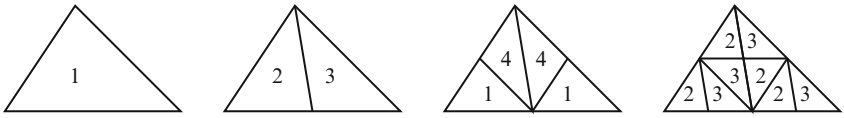


Fig. 1. Four similarity classes of triangles generated by the newest vertex bisection

The triangulation obtained by the newest vertex might have hanging nodes. We have to make additional subdivisions to eliminate the hanging nodes, i.e., complete the new partition. The completion should also follow the bisection rules. We shall consider more combinatory properties of the completion.

Let the triangles of the initial triangulation be assigned generation 0. We refer to the two triangles obtained by splitting a triangle τ in two sub-triangles by the newest vertex procedure as being the children of τ. For $i = 1, 2, \ldots$, we define the generation of the children of τ to be i if the parent τ has the generation $i-1$. It can be shown that the completion will terminate in finite steps, due to the fact that the completion process will not create new generations of triangles (see [3, 13]).

We ask more than the termination of the completion process. That is we want to control the number of elements refined due to the completion. To this end, we have to carefully assign the newest vertexs for the initial partition $\mathscr{T}_0$. A triangle is called *compatibly divisible* if its refinement edge is either the refinement edge of the triangle that shares that edge or an edge on the boundary. A triangulation $\mathscr{T}$ is called *compatibly divisible* or *compatibly labled* if every triangle is compatibly divisible. See Fig. 2 for an example of such compatible initial labeling.

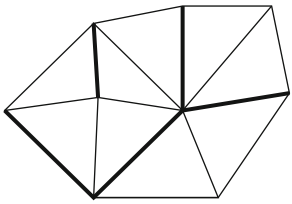


Fig. 2. A compatibly divisible labeling of the initial triangulation where edges in *bold case* are refinement edges

It is obvious that the completion for a compatible triangulation is terminated in one step. Mitchell [13] proves that for any conforming triangulation $\mathscr{T}$, there exist a compatible label scheme. Biedl et al. [4] present an $O(N)$ algorithm to find a compatible labeling for a triangulation $\mathscr{T}$ with N elements.

Let $\mathscr{T}_0$ be a compatible triangulation and let $\mathscr{T}_{\frac{1}{2}}$ be a triangulation obtained by the newest vertex bisection by performing m_0 bisections starting from $\mathscr{T}_0$. Denote by $\mathscr{M}_0$ the set of all m_0 marked and split triangles. Note that not all the triangles of $\mathscr{M}_0$ have to be in $\mathscr{T}_0$. Let $\mathscr{T}_1$ be the (minimal) conforming refinement of $\mathscr{T}_{\frac{1}{2}}$ and denote by n_k the number of triangles of $\mathscr{T}_k$, $k = 0, 1$ (Fig. 3).

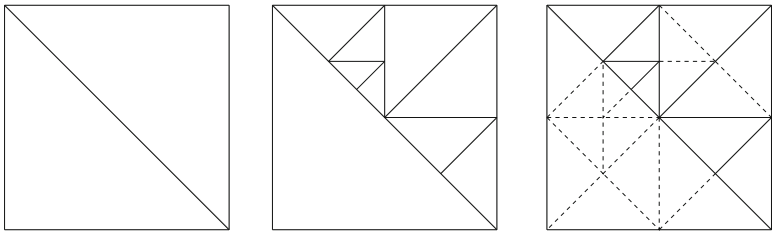


Fig. 3. Marking, splitting, and completing. **(a)** $\mathscr{T}_0$. **(b)** $\mathscr{T}_{\frac{1}{2}}$. **(c)** $\mathscr{T}_1$

Theorem 2. *Let $\mathscr{T}_0$ be a compatible triangulation and $\mathscr{T}_1$ be obtained as above. Then there exists a constant C only depending on the minimal angle of $\mathscr{T}_0$ such that*

$$n_1 \leq n_0 + (C+1)\, m_0. \tag{13}$$

Remark 1. It is a temptation to repeat the Theorem 2 to conclude: for $j = 1, 2, \ldots, p-1$, we have that $\mathscr{T}_{j+1}$ is obtained from $\mathscr{T}_j$, by m_j markings and then minimal completion, then

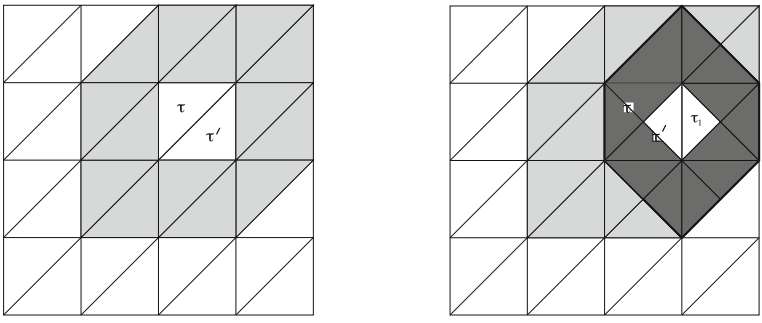
$$n_p \leq n_0 + (C+1)\,(m_0 + m_1 + \cdots + m_{p-1}). \tag{14}$$

Unfortunately this argument does not work since $\mathscr{T}_1$ may not be compatibly divisible anymore. The inequality (14) still holds but the proof is much involved; See Theorem 2.4 in [7]. The bound (13) can be derived from that theorem; See Lemma 2.5 in [7]. However, careful tracing the argument in [7] would give a huge constant in (14) in the magnitude of 10,000. We shall give another more direct and simpler proof based on an improved technique in [1]. The constant in our proof is much smaller and usually below 100. Note that numerically in the average case of the constant is around 4 and in the worst case is around 14; see [1].

Let us introduce notation for uniform bisection by setting $\overline{\mathscr{T}}_k$ as the triangulation obtained by bisecting each triangle in $\mathscr{T}_0$ completely up to the k-th generation. The assumption: $\mathscr{T}_0$ is compatible divisible implies that $\overline{\mathscr{T}}_k$ is conforming and compatible divisible for all $k \geq 1$. Note that this may not hold if the initial labeling is not compatibly divisible.

For a triangle τ, we define a neighbor of τ as another triangle sharing a common edges of τ. By the definition, a triangle has at most three neighbors. Among them, for $\tau \in \overline{\mathscr{T}}_k$, we define the *refinement neighbor* of τ as the triangle $\tau' \in \overline{\mathscr{T}}_k$ such that τ and τ' use the same edge as their refinement edges. We allow $\tau' = \varnothing$ for τ touching the boundary. We define the *barrier* of τ as all triangles in $\overline{\mathscr{T}}_{g(\tau)}$ which intersect $\tau \cup \tau'$ and denoted by $B(\tau)$, i.e.,

$$B(\tau) = \{\hat{\tau} \in \overline{\mathscr{T}}_{g(\tau)}, \hat{\tau} \cap (\tau \cup \tau') \neq \varnothing\}.$$



(a) (b)

Fig. 4. Barrier of a safe triangle. **(a)** Barrier 1. **(b)** Barrier 2

Definition 1. *We say that τ is a* safe *triangle if none of the barrier elements of τ is marked in going from $\mathscr{T}_0$ to $\mathscr{T}_1$, namely $\hat{\tau} \notin \mathscr{M}_0$ for any $\hat{\tau} \in B(\tau)$.*

The following lemma will justify the name of safe triangles. They are triangles that not touched going from $\mathscr{T}_0$ to $\mathscr{T}_1$.

Lemma 1. *Any safe triangle τ in $\mathscr{T}_0$ or born in the marking and completion process of going from $\mathscr{T}_0$ to $\mathscr{T}_1$ will never be bisected during the completion process.*

Proof. We shall prove it by the induction over the generation of τ. Suppose $g(\tau) = \max_{\tilde{\tau}\in\mathscr{T}_{\frac{1}{2}}} g(\tilde{\tau})$ and τ is safe. Then τ will not be bisected during the completion since the completion will not increase the maximal generation.

Assume that our statement holds for all safe triangles of generation $p+1$. We will show that the statement also holds for a safe triangle with generation p. Note that to trigger the bisection of τ, one has to refine one of the two neighbors of τ (which do not share the refinement edge with τ) twice or two such neighbors of τ' twice (since τ and τ' share the refinement edge). Without loss of generality, let us say that one of the neighbor τ' is bisected once in the completion process. Then it produces a children triangle τ_1 of generation $p+1$ which has a common edge with τ'. It is important to note that $B(\tau_1) \subset B(\tau)$ and thus τ_1 is safe; See Fig. 4 for an illustration. By the inductive hypothesis τ_1 will never be bisected anymore during the completion process. Consequently, τ will never be bisected during the completion process.

Now we are in the position to prove Theorem 2.

Proof. (of Theorem 2) We denote by $\mathscr{M}_{\frac{1}{2}}$ as the set of all triangles τ which are split in the completion process of going from $\mathscr{T}_{\frac{1}{2}}$ to $\mathscr{T}_1$. Let us choose a triangle $\tau \in \mathscr{M}_{\frac{1}{2}}$. Since τ is split in the completion process, by the above Lemma, τ is not safe. It implies that there should exist a same-generation triangle $F(\tau)$ in $B(\tau)$ such that $F(\tau) \in \mathscr{M}_0$. In this way, we defined a map from $F : \mathscr{M}_{\frac{1}{2}} \to \mathscr{M}_0$.

Note that F is not necessary a one-to-one map, but a triangle $\tau \in \mathscr{M}_0$ could be in only finite number of barriers, due to the space limitation of the same-generation assumption. Given a triangle τ, we define the first ring of τ as all triangles intersect τ and the second ring of τ as the union of first rings of triangles in the first ring of τ. Then τ can be only in the barrier of triangles in its second ring and thus the number is bounded by the maximum number of triangles in the second ring of a triangle, say C, which is usually below 100. Thus any triangle in $\mathscr{M}_0$ is the image of at most C triangles from $\mathscr{M}_{\frac{1}{2}}$. This leads to the fact that the number of splittings needed for completion can be bounded by Cm_0. Since any splitting in the completion process adds one more triangle towards the completed mesh $\mathscr{T}_1$, we have proved (13).

4 Local Grid Refinement Algorithm

In this section we shall propose a new approach for the local grid refinement based on the error estimate and the equidistribution principle. We will use newest vertex bisection to refine the grid and use $|u|_{2,1,\tau}$ as an error indicator. With a little bit higher regularity requirement of u, we are able to prove the effectiveness of our algorithm. Namely, it will end with an optimal asymptotic error estimate similar to (1).

4.1 Local Refinement Strategy

We will illustrate a way to find a nearly optimal grid for the solution of (2). We will use the newest vertex bisection refinement procedure with the marking strategy given by (11). For the later analysis, we will have to assume that the solution u is in $W^{2,1}$ and that the Hardy-Littlewood maximal function of D^2u is in $L^1(\Omega)$. Due to a result of [17], this is equivalently $D^2u \in L\log L(\Omega)$. Such further assumption holds if for example $u \in W^{2,p}$ for some $p > 1$.

The maximal function of an integrable function f on Ω is defined by

$$\widetilde{M}f(x) = \sup \frac{1}{|Q|} \int_Q |f(y)|\, dy,$$

where the supremum is taken over all square domains contained in Ω and containing x.

For a triangulation obtained by the newest vertices bisection from $\mathscr{T}_0$. The similarity classes are in fact completely represented by the children and grandchildren of all triangles from $\mathscr{T}_0$. Let us denote by $\mathscr{C}_0$ the following family of triangles:

$$\mathscr{C}_0 = \{\tau |\ \tau \text{ is a triangle contained in } \Omega \text{ and is similar with a child or grandchild of a triangle from } \mathscr{T}_0\}$$

We define another maximal function

$$Mf(x) = \sup \frac{1}{|\tau|} \int_\tau |f(y)|\, dy,$$

where the supremum is taken over all triangles $\tau \in \mathscr{C}_0$ containing x. Then it is easy to show that $\widetilde{M}$ and M are equivalent in the sense that

$$c_1\widetilde{M}f(x) \le Mf(x) \le c_2\widetilde{M}f(x), \quad \forall x \in \Omega$$

with c_1 and c_2 independent of x. Thus, for theoretical purposes, the two operators M and $\widetilde{M}$ are interchangeable.

The following result concerns the number of the new triangles added in the refinement procedure. The main idea of the proof for the 1-D case was showed to the authors by DeVore and can be found in [10].

Theorem 3. *Let f be an integrable function on Ω such that $Mf \in L^1(\Omega)$, and let $\varepsilon > 0$ be given. Assume that the newest vertex bisection refinement procedure is applied to an compatible initial triangulation $\mathscr{T}_0$ with n_0 triangles. Let the marking strategy be given by: a triangle τ is marked if*

$$\int_\tau |f(x)|\, dx > \varepsilon.$$

Denote by $\mathscr{M}_0$ the set of all marked and split triangles. Then, the marking and refinement procedure will terminate in finite steps and we have

$$n_0 + m_0 < \frac{2}{\varepsilon} \int_\Omega Mf(x)\, dx, \tag{15}$$

where m_0 is the number of elements of $\mathscr{M}_0$. Assume that $\mathscr{T}_{\frac{1}{2}}$ is the triangulation obtained from $\mathscr{T}_0$ after the m_0 bisections. Let $\mathscr{T}_1$ be the (minimal) conforming refinement of $\mathscr{T}_{\frac{1}{2}}$ and denote by n_1 the number of triangles of $\mathscr{T}_1$. Then,

$$n_1 \leq \frac{C_1}{\varepsilon} \int_\Omega |Mf(x)|\, dx, \tag{16}$$

with a constant C_1 independent of the function f and the number ε. More precisely, $C_1 = 2(C+1)$, with C the constant of Theorem 2.

Proof. Since $\lim_{|\tau|\to 0} \int_\tau |f(x)|\, dx = 0$ and the areas of new triangles are exponentially decreased, the refinement procedure will terminate in finite steps.

We can assume without loss of generality that each triangle in $\mathscr{T}_{\frac{1}{2}}$ is not a triangle in $\mathscr{T}_0$. Now, let $\tau \in \mathscr{T}_{\frac{1}{2}}$ and let $\tilde{\tau}$ be its parent. Then $\tilde{\tau} \in \mathscr{M}_0$. (Recall that $\mathscr{M}_0$ is the collection of marked triangles in the refinement procedure.) By our refinement strategy

$$\int_{\tilde{\tau}} |f(x)|\, dx > \varepsilon,$$

Thus,

$$Mf(x) > \frac{1}{|\tilde{\tau}|} \int_{\tilde{\tau}} |f(y)|\, dy > \frac{\varepsilon}{|\tilde{\tau}|}, \quad \forall x \in \tau.$$

Integrating the above inequality on τ we have,

$$\int_\tau Mf(x)\, dx > \frac{\varepsilon}{2}. \tag{17}$$

Here we use fact $|\tilde{\tau}| = 2|\tau|$. If we sum up (17) over all $n_0 + m_0$ triangles $\tau \in \mathscr{T}_{\frac{1}{2}}$ we obtain (15).

By using Theorem 2 we have that

$$n_1 \leq n_0 + m_0 + C\, m_0 \leq (C+1)\, (n_0 + m_0).$$

The estimate (16) follows now as a direct consequence of (15) and the above inequality.

An application of Theorem 1 and the estimate (16) for $f = D^2 u$ and $\varepsilon = 1/N$, leads to the proof of the existence of a nearly optimal grid. Starting from a coarse grid $\mathscr{T}_0$, we define the approximation class $\mathscr{A}_{1/2}$ as

$$\mathscr{A}_{1/2} = \{u \in H_0^1(\Omega) : |u|_{\mathscr{A}_{1/2}} := \sup_{N \geq \#\mathscr{T}_0} N^{-1/2} \inf_{\#\mathscr{T} \leq N} \inf_{v_h \in V(\mathscr{T})} |u - v_h|_1 < \infty\}.$$

Corollary 1. *If $u \in W^{2,L\log L}(\Omega)$, then $u \in \mathscr{A}_{1/2}$.*

Remark 2. The $(L\log L)$ norm is needed only for proving the success of the algorithm but is not effectively needed for the implementation of the algorithm. If we can find good approximations or upper bound for $\int_\tau D^2 u\, dx$ on triangles using e.g., gradient and Hessian recovery methods (from the discrete Galerkin approximation of u) or using regularity result in [12], then the ideas presented in this paper can lead to new and optimal adaptive methods.

Bibliography

[1] F. B. Atalay and D. M. Mount. The cost of compatible refinement of simplex decomposition trees. In *Proceedings of the 15th International Meshing Roundtable, Birmingham, AL, September 2006*, pages 57–70, 2006.

[2] I. Babuška and W. C. Rheinboldt. A posteriori error estimates for the finite element method. *Internat. J. Numer. Methods Engrg.*, 12:1597–1615, 1978.

[3] E. Bänsch. Local mesh refinement in 2 and 3 dimensions. *Impact Comput. Sci. Engrg.*, 3:181–191, 1991.

[4] T. C. Biedl, P. Bose, E. D. Demaine, and A. Lubiw. Efficient algorithms for Petersen's matching theorem. *J. Algorithms*, 38(1):110–134, 2001. ISSN 0196-6774. doi: http://dx.doi.org/10.1006/jagm.2000.1132.

[5] P. Binev and R. DeVore. Fast computation in adaptive tree approximation. *Numer. Math.*, 97:193–217, 2004.

[6] P. Binev, W. Dahmen, R. DeVore, and P. Petrushev. Approximation classes for adaptive methods. *Serdica Math. J*, 28:391–416, 2002.

[7] P. Binev, W. Dahmen, and R. DeVore. Adaptive finite element methods with convergence rates. *Numer. Math.*, 97(2):219–268, 2004.

[8] J. M. Cascón, C. Kreuzer, R. H. Nochetto, and K. G. Siebert. Quasi-optimal convergence rate for an adaptive finite element method. *SIAM J. Numer. Anal.*, 46(5):2524–2550, 2008.

[9] L. Chen and J. Xu. Convergence of adaptive finite element methods. In T. Tang and J. Xu, editors, *Adaptive Computations: Theory and Algorithms*, pages 9–37. Science Press, Beijing, 2007.

[10] R. A. DeVore. Nonlinear approximation. *Acta Numer.*, pages 51–150, 1998.

[11] W. Dörfler. A convergent adaptive algorithm for Poisson's equation. *SIAM J. Numer. Anal.*, 33:1106–1124, 1996.

[12] P. Grisvard. *Elliptic Problems in Nonsmooth Domains*. Pitman, Boston, 1995.

[13] W. F. Mitchell. *Unified Multilevel Adaptive Finite Element Methods for Elliptic Problems*. PhD thesis, University of Illinois at Urbana-Champaign, 1988.

[14] W. F. Mitchell. A comparison of adaptive refinement techniques for elliptic problems. *ACM Trans. Math. Softw. (TOMS) archive*, 15(4):326–347, 1989.

[15] P. Morin, R. H. Nochetto, and K. G. Siebert. Data oscillation and convergence of adaptive FEM. *SIAM J. Numer. Anal.*, 38(2):466–488, 2000.

[16] P. Morin, R. H. Nochetto, and K. G. Siebert. Convergence of adaptive finite element methods. *SIAM Rev.*, 44(4):631–658, 2002.

[17] R. M. Stein. Note on the class L(logL). *Studia. Math.*, 32:305–310, 1969.

[18] R. Stevenson. Optimality of a standard adaptive finite element method. *Found. Comput. Math.*, 7(2):245–269, 2007.
[19] R. Verfürth. *A Review of A Posteriori Error Estimation and Adaptive Mesh Refinement Tecniques*. B. G. Teubner, 1996.

Some Recent Tools and a BDDC Algorithm for 3D Problems in H(curl)

Clark R. Dohrmann[1] and Olof B. Widlund[2]

[1] Sandia National Laboratories, Albuquerque, New Mexico, 87185-0346, USA. Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000, crdohrm@sandia.gov

[2] Courant Institute, 251 Mercer Street, New York, NY 10012, USA. This work supported in part by the U.S. Department of Energy under contracts DE-FG02-06ER25718 and in part by National Science Foundation Grant DMS-0914954, widlund@cims.nyu.edu, http://www.cs.nyu.edu/cs/faculty/widlund

Summary. We present some recent domain decomposition tools and a BDDC algorithm for 3D problems in the space $H(\mathrm{curl};\Omega)$. Of primary interest is a face decomposition lemma which allows us to obtain improved estimates for a BDDC algorithm under less restrictive assumptions than have appeared previously in the literature. Numerical results are also presented to confirm the theory and to provide additional insights.

1 Introduction

We investigate a BDDC algorithm for three-dimensional (3D) problems in the space $H_0(\mathrm{curl};\Omega)$. The subject problem is to obtain edge finite element approximations of the variational problem: Find $\boldsymbol{u} \in H_0(\mathrm{curl};\Omega)$ such that

$$a_\Omega(\boldsymbol{u},\boldsymbol{v}) = (\boldsymbol{f},\boldsymbol{v})_\Omega \quad \forall \boldsymbol{v} \in H_0(\mathrm{curl};\Omega),$$

where

$$a_\Omega(\boldsymbol{u},\boldsymbol{v}) := \int_\Omega [(\alpha \nabla \times \boldsymbol{u} \cdot \nabla \times \boldsymbol{v}) + (\beta \boldsymbol{u} \cdot \boldsymbol{v})]\, dx, \quad (\boldsymbol{f},\boldsymbol{v})_\Omega = \int_\Omega \boldsymbol{f} \cdot \boldsymbol{v}\, dx.$$

The norm of $\boldsymbol{u} \in H(\mathrm{curl};\Omega)$, for a domain with diameter 1, is given by $a_\Omega(\boldsymbol{u},\boldsymbol{u})^{1/2}$ with $\alpha = 1$ and $\beta = 1$; the elements of $H_0(\mathrm{curl})$ have vanishing tangential components on $\partial\Omega$. We could equally well consider cases where this boundary condition is imposed only on one or several subdomain faces which form part of $\partial\Omega$. We will assume that $\alpha \geq 0$ and $\beta > 0$ are constant in each of the subdomains $\Omega_1, \ldots, \Omega_N$. Our results could be presented in a form which accommodates properties which are not constant or isotropic in each subdomain, but we avoid this generalization for purposes of clarity.

R. Bank et al. (eds.), *Domain Decomposition Methods in Science and Engineering XX*, Lecture Notes in Computational Science and Engineering 91,
DOI 10.1007/978-3-642-35275-1_2,

In the pioneering work of [12], two different cases were analyzed for FETI-DP algorithms:
Case 1:

$$\alpha_i = \alpha \quad \text{for} \quad i = 1, \ldots, N$$

The condition number bound reported for the preconditioned operator is

$$\kappa \leq C \max_i (1 + H_i^2 \beta_i / \alpha)(1 + \log(H/h))^4, \tag{1}$$

where $H/h := \max_i H_i/h_i$.
Case 2:

$$\beta_i = \beta \quad \text{for} \quad i = 1, \ldots, N$$

for which the reported condition number bound is

$$\kappa \leq C \max_i (1 + H_i^2 \beta / \alpha_i)(1 + \log(H/h))^4. \tag{2}$$

We address the following basic questions regarding [12] in this study.

1. Is is possible to remove the assumption of $\alpha_i = \alpha$ or $\beta_i = \beta$ for all i?
2. Is it possible to remove the factor of $H_i^2 \beta_i / \alpha_i$ from the estimates?
3. Is is possible to reduce the logarithmic factor from four powers to two powers as is typical of other iterative substructuring algorithms?
4. Do FETI-DP or BDDC algorithms for 3D H(curl) problems have certain complications not present for problems with just a single parameter?

We find in the following sections that the answers are yes to all four questions. However, due to page limitations, we only consider here the relatively rich coarse space of Algorithm C of [12]. We remark that the analysis of 3D H(curl) problems with material property jumps between subdomains is quite limited in the literature. A comprehensive treatment of problems in 2D can be found in [3]. A different iterative substructuring algorithm for 3D problems is given in [6], but the authors were unable to conclude whether their condition number bound was independent of material property jumps. A related study on substructuring preconditioners can also be found in [7].

2 Tools

We assume that Ω is decomposed into N non-overlapping subdomains, $\Omega_1, \ldots, \Omega_N$, each the union of elements of the triangulation of Ω. We denote by H_i the diameter of Ω_i. The interface of the domain decomposition is given by

$$\Gamma := \left(\bigcup_{i=1}^{N} \partial \Omega_i \right) \setminus \partial \Omega,$$

and the contribution to Γ from $\partial\Omega_i$ by $\Gamma_i := \partial\Omega_i \backslash \partial\Omega$. These sets are unions of subdomain faces, edges, and vertices. For simplicity, we assume that each subdomain is a shape-regular and convex tetrahedron or hexahedron with planar faces.

We assume a shape-regular triangulation $\mathscr{T}_{h_i}$ of each Ω_i with nodes matching across the interfaces. The smallest element diameter of $\mathscr{T}_{h_i}$ is denoted by h_i. Associated with the triangulation $\mathscr{T}_{h_i}$ are the two finite element spaces $W^{h_i}_{\text{grad}} \subset H(\text{grad}, \Omega_i)$ and $W^{h_i}_{\text{curl}} \subset H(\text{curl}, \Omega_i)$ based on continuous, piecewise linear, tetrahedral nodal elements and linear, tetrahedral edge (Nédeléc) elements, respectively. We could equally well develop our algorithms and theory for low order hexahedral elements.

The energy of a vector function $\boldsymbol{u} \in W^{h_i}_{\text{curl}}$ for subdomain Ω_i is defined as

$$E_i(\boldsymbol{u}) := \alpha_i(\nabla \times \boldsymbol{u}, \nabla \times \boldsymbol{u})_{\Omega_i} + \beta_i(\boldsymbol{u}, \boldsymbol{u})_{\Omega_i}, \tag{3}$$

where α_i and β_i are assumed constant in Ω_i.

Let $\boldsymbol{N}_e \in W^{h_i}_{\text{curl}}$ and $\boldsymbol{t}_e$ denote the finite element shape function and unit tangent vector, respectively, for an edge e of $\mathscr{T}_{h_i}$. We assume that $\boldsymbol{N}_e$ is scaled such that $\boldsymbol{N}_e \cdot \boldsymbol{t}_e = 1$ along e. The *edge* finite element interpolant of a sufficiently smooth vector function $\boldsymbol{u} \in H(\text{curl}, \Omega_i)$ is then defined as

$$\Pi^{h_i}(\boldsymbol{u}) := \sum_{e \in \mathscr{M}_{\bar{\Omega}_i}} u_e \boldsymbol{N}_e, \quad u_e := (1/|e|) \int_e \boldsymbol{u} \cdot \boldsymbol{t}_e \, ds, \tag{4}$$

where $\mathscr{M}_{\bar{\Omega}_i}$ is the set of edges of $\mathscr{T}_{h_i}$, and $|e|$ is the length of e. We will also make use of other sets of edges of $\mathscr{T}_{h_i}$. Namely, $\mathscr{M}_{\partial\Omega_i}$, $\mathscr{M}_{\mathscr{E}}$, $\mathscr{M}_{\mathscr{F}}$, and $\mathscr{M}_{\partial\mathscr{F}}$ contain the edges of $\partial\Omega_i$, subdomain edge $\mathscr{E}$, subdomain face $\mathscr{F}$, and $\partial\mathscr{F}$, respectively. We denote by $\mathscr{G}_{i\mathscr{F}}$, $\mathscr{G}_{i\mathscr{E}}$, and $\mathscr{G}_{i\mathscr{V}}$ sets of subdomain faces, subdomain edges, and subdomain vertices for Ω_i. The wire basket $\mathscr{W}_i$ is the union of all subdomain edges and vertices for Ω_i. We will also make use of the symbol $\omega_i := 1 + \log(H_i/h_i)$, and bold faced symbols refer to vector functions. We denote by $\bar{p}_i$ the mean of p_i over Ω_i.

The estimate in the next lemma can be found in several references, see e.g., Lemma 4.16 of [13].

Lemma 1. *For any $p_i \in W^{h_i}_{grad}$ and subdomain edge $\mathscr{E}$ of Ω_i,*

$$\|p_i\|^2_{L^2(\mathscr{E})} \leq C\omega_i \|p_i\|^2_{H^1(\Omega_i)}. \tag{5}$$

Lemma 2. *For any $p_i \in W^{h_i}_{grad}$, there exist $p_{i\mathscr{V}}, p_{i\mathscr{E}}, p_{i\mathscr{F}} \in W^{h_i}_{grad}$ such that*

$$p_i|_{\partial\Omega_i} = \sum_{\mathscr{V} \in \mathscr{G}_{i\mathscr{V}}} p_{i\mathscr{V}}|_{\partial\Omega_i} + \sum_{\mathscr{E} \in \mathscr{G}_{i\mathscr{E}}} p_{i\mathscr{E}}|_{\partial\Omega_i} + \sum_{\mathscr{F} \in \mathscr{G}_{i\mathscr{F}}} p_{i\mathscr{F}}|_{\partial\Omega_i}, \tag{6}$$

where the nodal values of $p_{i\mathscr{V}}$, $p_{i\mathscr{E}}$, and $p_{i\mathscr{F}}$ on $\partial\Omega_i$ may be nonzero only at the nodes of $\mathscr{V}$, $\mathscr{E}$, and $\mathscr{F}$, respectively. Further,

$$|p_{i\mathscr{V}}|^2_{H^1(\Omega_i)} \leq C\|p_i\|^2_{H^1(\Omega_i)}, \tag{7}$$

$$|p_{i\mathscr{E}}|^2_{H^1(\Omega_i)} \leq C\omega_i\|p_i\|^2_{H^1(\Omega_i)}, \tag{8}$$

$$|p_{i\mathscr{F}}|^2_{H^1(\Omega_i)} \leq C\omega_i^2\|p_i\|^2_{H^1(\Omega_i)}. \tag{9}$$

Proof. The estimates in (7)–(9) are standard, and follow from Corollary 4.20 and Lemma 4.24 of [13] and elementary estimates.

We note that a Poincaré inequality allows us to replace the H^1-norm of p_i by its H^1-seminorm in Lemmas 1 and 2 if $\bar{p}_i = 0$.

The next lemma is stated without proof due to page restrictions.

Lemma 3. *Let $f_i \in W^{h_i}_{grad}$ have vanishing nodal values everywhere on $\partial\Omega_i$ except on the wire basket $\mathscr{W}_i$ of Ω_i. For each subdomain face $\mathscr{F}$ of Ω_i and $Ch_i \leq d \leq H_i/C$, $C > 1$, there exists a $\boldsymbol{v}_i \in W^{h_i}_{curl}$ such that $v_{ie} = \nabla f_{ie}$ for all $e \in \mathscr{M}_{\mathscr{F}}$, $v_{ie} = 0$ for all other edges of $\partial\Omega_i$, and*

$$\|\boldsymbol{v}_i\|^2_{L^2(\Omega_i)} \leq C(\omega_i\|f_i\|^2_{L^2(\partial\mathscr{F})} + d^2\|\nabla f_i \cdot \boldsymbol{t}_{\partial\mathscr{F}}\|^2_{L^2(\partial\mathscr{F})}), \tag{10}$$

$$\|\nabla \times \boldsymbol{v}_i\|^2_{L^2(\Omega_i)} \leq C(\tau(d)\|f_i\|^2_{L^2(\partial\mathscr{F})} + \|\nabla f_i \cdot \boldsymbol{t}_{\partial\mathscr{F}}\|^2_{L^2(\partial\mathscr{F})}), \tag{11}$$

where $\boldsymbol{t}_{\partial\mathscr{F}}$ is a unit tangent along $\partial\mathscr{F}$, and

$$\tau(d) = \begin{cases} 0 & \text{if } d > H_i/C \\ d^{-2} & \text{otherwise.} \end{cases}$$

The Helmholtz-type decomposition and estimates in the next lemma will allow us to make use of and build on existing tools for scalar functions in $H^1(\Omega_i)$. We refer the reader to Lemma 5.2 of [4] for the case of convex polyhedral subdomains; this important paper was preceded by Hiptmair et al. [5], which concerns other applications of the same decomposition.

Lemma 4. *For a convex and polyhedral subdomain Ω_i and any $\boldsymbol{u}_i \in W^{h_i}_{curl}$, there is a $\boldsymbol{q}_i \in W^{h_i}_{curl}$, $\boldsymbol{\Psi}_i \in (W^{h_i}_{grad})^3$, and $p_i \in W^{h_i}_{grad}$ such that*

$$\boldsymbol{u}_i = \boldsymbol{q}_i + \Pi^{h_i}(\boldsymbol{\Psi}_i) + \nabla p_i, \tag{12}$$

$$\|\nabla p_i\|_{L^2(\Omega_i)} \leq C\|\boldsymbol{u}_i\|_{L^2(\Omega_i)}, \tag{13}$$

$$\|\boldsymbol{\Psi}_i\|_{L^2(\Omega_i)} \leq C\|\boldsymbol{u}_i\|_{L^2(\Omega_i)}, \tag{14}$$

$$\|h_i^{-1}\boldsymbol{q}_i\|^2_{L^2(\Omega_i)} + \|\boldsymbol{\Psi}_i\|^2_{H^1(\Omega_i)} \leq C\|\nabla \times \boldsymbol{u}_i\|^2_{L^2(\Omega_i)}. \tag{15}$$

Lemma 5. *For any $\boldsymbol{u}_i \in W_{curl}^{h_i}$ with $u_{ie} = 0$ for all $e \in \mathscr{M}_{\partial\mathscr{F}}$, there exists a $\boldsymbol{v}_{i\mathscr{F}} \in W_{curl}^{h_i}$ such that $v_{i\mathscr{F}e} = u_{ie}$ for all $e \in \mathscr{M}_{\mathscr{F}}$, $v_{i\mathscr{F}e} = 0$ for all $e \in \mathscr{M}_{\partial\Omega_i} \setminus \mathscr{M}_{\mathscr{F}}$, and*

$$E_i(\boldsymbol{v}_{i\mathscr{F}}) \leq C\omega_i^2 E_i(\boldsymbol{u}_i), \tag{16}$$

where the energy E_i is defined in (3).

Proof. Let p_i in (12) be chosen so $\bar{p}_i = 0$. This is possible since a constant can be added to p_i without changing its gradient. Because $u_{ie} = 0$ for all $e \in \mathscr{M}_{\partial\mathscr{F}}$, it follows from Lemmas 1 and 4 and elementary estimates that

$$\begin{aligned}\|\nabla p_i \cdot \boldsymbol{t}_{\mathscr{E}}\|^2_{L^2(\partial\mathscr{F})} &= \|(\Pi^{h_i}(\boldsymbol{\Psi_i}) + \boldsymbol{q}_i) \cdot \boldsymbol{t}_{\mathscr{E}}\|^2_{L^2(\partial\mathscr{F})} \\ &\leq C\omega_i \|\nabla \times \boldsymbol{u}_i\|^2_{L^2(\Omega_i)}.\end{aligned} \tag{17}$$

We then find from Lemmas 2 and 4 that

$$\|\nabla p_{i\mathscr{F}}\|^2_{L^2(\Omega_i)} \leq C\omega_i^2 \|\boldsymbol{u}_i\|^2_{L^2(\Omega_i)}. \tag{18}$$

Define

$$p_{i\mathscr{W}} := \sum_{\mathscr{V} \in \mathscr{G}_{i\mathscr{V}}} p_{i\mathscr{V}} + \sum_{\mathscr{E} \in \mathscr{G}_{i\mathscr{E}}} p_{i\mathscr{E}}, \quad d := \begin{cases} H_i & \text{if } d_i \geq H_i \\ \max(d_i, Ch_i) & \text{otherwise,} \end{cases}$$

where $d_i := \sqrt{\alpha_i/\beta_i}$. Further, let $p_{i\mathscr{W}}$ and $\boldsymbol{p}_{i\mathscr{F}}$ denote the functions f_i and $\boldsymbol{v}_i$, respectively, of Lemma 3. We then find from Lemmas 1 and 3 and (17) that

$$E_i(\boldsymbol{p}_{i\mathscr{F}}) \leq C\omega_i^2 E_i(\boldsymbol{u}_i), \tag{19}$$

where $p_{i\mathscr{F}e} = \nabla p_{i\mathscr{W}e}\ \forall e \in \mathscr{M}_{\mathscr{F}}$ and $p_{i\mathscr{F}e} = 0\ \forall e \in \mathscr{M}_{\partial\Omega_i} \setminus \mathscr{M}_{\mathscr{F}}$. With reference to (12) and (4), we define

$$\boldsymbol{q}_{i\mathscr{F}} := \sum_{e \in \mathscr{M}_{\mathscr{F}}} q_{ie} \boldsymbol{N}_e, \tag{20}$$

and from elementary finite element estimates and Lemma 4 find

$$\|\boldsymbol{q}_{i\mathscr{F}}\|^2_{L^2(\Omega_i)} \leq Ch_i^3 \sum_{e \in \mathscr{M}_{\mathscr{F}}} q_{ie}^2 \leq C\|\boldsymbol{q}_i\|^2_{L^2(\Omega_i)} \leq C\|\boldsymbol{u}_i\|^2_{L^2(\Omega_i)}, \tag{21}$$

$$\|\nabla \times \boldsymbol{q}_{i\mathscr{F}}\|^2_{L^2(\Omega_i)} \leq Ch_i \sum_{e \in \mathscr{M}_{\mathscr{F}}} q_{ie}^2 \leq C\|\nabla \times \boldsymbol{u}_i\|^2_{L^2(\Omega_i)}. \tag{22}$$

It follows from Lemmas 2 and 4 that there exists a $\boldsymbol{\Psi}_{i\mathscr{F}} \in (W_{\text{grad}}^{h_i})^3$ such that $\boldsymbol{\Psi}_{i\mathscr{F}} = \boldsymbol{\Psi}_i$ at all nodes of $\mathscr{F}$, that vanishes at all other nodes of $\partial\Omega_i$, and

$$\|\boldsymbol{\Psi}_{i\mathscr{F}}\|^2_{L^2(\Omega_i)} \leq C\|\boldsymbol{\Psi}_i\|^2_{L^2(\Omega_i)} \leq C\|\boldsymbol{u}_i\|^2_{L^2(\Omega_i)}, \tag{23}$$

$$\|\nabla \times \boldsymbol{\Psi}_{i\mathscr{F}}\|^2_{L^2(\Omega_i)} \leq C\omega_i^2 \|\boldsymbol{\Psi}_i\|^2_{H^1(\Omega_i)} \leq C\omega_i^2 \|\nabla \times \boldsymbol{u}_i\|^2_{L^2(\Omega_i)}. \tag{24}$$

From Lemmas 1 and 4, we obtain

$$\|\boldsymbol{\Psi}_i\|^2_{L^2(\partial\mathscr{F})} \leq C\omega_i \|\boldsymbol{\Psi}_i\|^2_{H^1(\Omega_i)} \leq C\omega_i \|\nabla \times \boldsymbol{u}_i\|^2_{L^2(\Omega_i)}. \tag{25}$$

Let $\boldsymbol{\Psi}_{i\partial\mathscr{F}} \in (W^{h_i}_{\text{grad}})^3$ be identical to $\boldsymbol{\Psi}_i$ at all nodes of $\partial\mathscr{F}$ and vanish at all other nodes of Ω_i. For $\boldsymbol{g} := \Pi^{h_i}(\boldsymbol{\Psi}_{i\partial\mathscr{F}})$, we define

$$\boldsymbol{g}_{i\mathscr{F}} := \sum_{e\in\mathscr{M}_\mathscr{F}} g^{h_i}_e \boldsymbol{N}_e. \tag{26}$$

From elementary estimates and (25), we then obtain

$$\|\boldsymbol{g}_{i\mathscr{F}}\|^2_{L^2(\Omega_i)} \leq Ch_i^2 |\boldsymbol{\Psi}_i|^2_{L^2(\partial\mathscr{F})} \leq C\omega_i h_i^2 \|\nabla \times \boldsymbol{u}_i\|^2_{L^2(\Omega_i)}, \tag{27}$$

$$\|\nabla \times \boldsymbol{g}_{i\mathscr{F}}\|^2_{L^2(\Omega_i)} \leq C\omega_i \|\nabla \times \boldsymbol{u}_i\|^2_{L^2(\Omega_i)}. \tag{28}$$

Defining

$$\boldsymbol{v}_{i\mathscr{F}} := \nabla p_{i\mathscr{F}} + \boldsymbol{p}_{i\mathscr{F}} + \boldsymbol{q}_{i\mathscr{F}} + \Pi^{h_i}(\boldsymbol{\Psi}_{i\mathscr{F}}) + \boldsymbol{g}_{i\mathscr{F}}, \tag{29}$$

we find that $v_{i\mathscr{F}e} = u_{ie}\ \forall e \in \mathscr{M}_\mathscr{F}$ and $v_{i\mathscr{F}e} = 0\ \forall e \in \mathscr{M}_{\partial\Omega_i} \setminus \mathscr{M}_\mathscr{F}$. The estimate in (16) then follows from the bounds for each of the terms on the right-hand-side of (29) along with elementary estimates for $\Pi^{h_i}(\boldsymbol{\Psi}_{i\mathscr{F}})$. □

3 BDDC

Background information and related theory for BDDC can be found in several references including [1, 2, 9–11]. Let u_i and u denote vectors of finite element coefficients associated with Γ_i and Γ. In general, entries in u_i and u_j are allowed to differ for $j \neq i$ even though they refer to the same finite element edge. Entries in the vector $\tilde{u}_i$ are partially continuous in the sense that specific edge values or edge averages over certain subsets of Γ are required to match for adjacent subdomains. In order to obtain consistent entries, we define the weighted average

$$\hat{u}_i = R_i \sum_{j=1}^{N} R_j^T D_j \tilde{u}_j, \tag{30}$$

where R_j is a 0–1 (Boolean) matrix that selects the rows of u_j from u and D_j is a weight matrix. The weight matrices form a partition of unity in the sense that

$$\sum_{i=1}^{N} R_i^T D_i R_i = I, \tag{31}$$

where I is the identity matrix. To summarize, $\hat{u}_i$ is fully continuous while $\tilde{u}_i$ is only partially continuous. The number of continuity constraints that must be satisfied by all the $\tilde{u}_i$ determines the dimension of the coarse space.

The energy of $\boldsymbol{u}$ for Ω_i can be expressed as

$$E_i(\boldsymbol{u}) = E_i(u_i) = u_i^T S_i u_i, \tag{32}$$

where S_i is the Schur complement matrix associated with Ω_i and Γ_i. The system operator for BDDC is the assembled Schur complement

$$S = \sum_{i=1}^{N} R_i^T S_i R_i. \tag{33}$$

From Theorem 25 of [11], the condition number of the BDDC preconditioned operator is bounded above by

$$\kappa(M^{-1}S) \leq \sup_{\tilde{u}_i} \frac{\sum_{i=1}^{N} \hat{u}_i^T S_i \hat{u}_i}{\sum_{i=1}^{N} \tilde{u}_i^T S_i \tilde{u}_i}. \tag{34}$$

This remarkably simple expression shows that the continuity constraints for $\tilde{u}_i$ should be chosen so that large increases in energy do not result from the averaging operation in (30).

Let $R_{i\partial\mathscr{F}_{ij}}$ select the rows of u_i corresponding to the edge coefficients on the boundary of the face $\mathscr{F}_{ij}$, the closure of which is $\partial\Omega_i \cap \partial\Omega_j$. Similarly, let $R_{i\mathscr{F}_{ij}}$ select the rows of u_i corresponding to the interior of the face $\mathscr{F}_{ij}$. We define the vector of face edge coefficients by $u_{iF} := R_{i\mathscr{F}_{ij}} u_i$ and the face Schur complement matrix by $S_{iFF} := R_{i\mathscr{F}_{ij}} S_i R_{i\mathscr{F}_{ij}}^T$.

Because of page restrictions, we only consider a very rich coarse space which includes every edge variable of each subdomain edge. This coarse space corresponds to Algorithm C of [12]. For this case, we choose the weighted average of u_{iF} and u_{jF} as

$$\hat{u}_F = (S_{iFF} + S_{jFF})^{-1}(S_{iFF} u_{iF} + S_{jFF} u_{jF}). \tag{35}$$

Thus,

$$u_{iF} - \hat{u}_F = (S_{iFF} + S_{jFF})^{-1} S_{jFF}(u_{iF} - u_{jF}). \tag{36}$$

Using the eigenvectors of the generalized eigenvalue problem $S_{iFF}x = \lambda S_{jFF}x$ as a convenient basis, we find

$$u_{kF}^T \bar{S}_{iFF} u_{kF} \leq u_{kF}^T S_{kFF} u_{kF}, \quad \forall u_{kF} \quad k \in \{i, j\}, \tag{37}$$

where

$$\bar{S}_{iFF} := S_{jFF}(S_{iFF} + S_{jFF})^{-1} S_{iFF} (S_{iFF} + S_{jFF})^{-1} S_{jFF} \tag{38}$$

Let us assume for the moment that there are vectors u_{ij}, u_{ji}, and a scalar $\hat{C} > 0$ such that

$$R_{i\partial\mathscr{F}_{ij}} u_{ij} = R_{j\partial\mathscr{F}_{ij}} u_{ji} = u_{\partial F}, \tag{39}$$

$$R_{i\mathscr{F}_{ij}} u_{ij} = R_{j\mathscr{F}_{ij}} u_{ji}, \tag{40}$$

$$u_{ij}^T S_i u_{ij} + u_{ji}^T S_j u_{ji} \leq \hat{C}(u_i^T S_i u_i + u_j^T S_j u_j). \tag{41}$$

In other words, u_{ij}, u_{ji}, u_i and u_j are all identical along the boundary of $\mathscr{F}_{ij}$. Further, u_{ij} and u_{ji} are identical in the interior of $\mathscr{F}_{ij}$, and the sum of their energies is bounded uniformly by the sum of the energies of u_i and u_j.

In order to establish a condition number bound for Algorithm C, we need an estimate for $E_i(R^T_{i\mathscr{F}_{ij}}(u_{iF}-\hat{u}_F))$; see (34). By construction, we have $R_{i\partial\mathscr{F}_{ij}}(u_i-u_{ij})=0$ and $R_{j\partial\mathscr{F}_{ij}}(u_j-u_{ji})=0$. Since $u_{iF}-u_{jF}=(u_{iF}-u_{ijF})-(u_{jF}-u_{jiF})$, it then follows from (36), (37), (41), and Lemma 5 that

$$\begin{aligned} E_i(R^T_{iF_{ij}}(u_{iF}-\hat{u}_F)) =& E_i(R^T_{iF_{ij}}(S_{iFF}+S_{jFF})^{-1}S_{jFF}(u_{iF}-u_{jF})) \\ \leq& 2(u_{iF}-u_{ijF})^T S_{iFF}(u_{iF}-u_{ijF})+ \\ & 2(u_{jF}-u_{jiF})^T S_{jFF}(u_{jF}-u_{jiF}) \\ \leq& \hat{C}C\omega_i^2(E_i(u_i)+E_j(u_j)). \end{aligned} \tag{42}$$

We are able to show there exist u_{ij} and u_{ji} which satisfy the conditions in (39)–(41) with $\hat{C}$ independent of mesh parameters and the material properties α_i, β_i, α_j, and β_j under the assumption

$$\alpha_m \leq C\alpha_n \quad \text{and} \quad \beta_m \leq C\beta_n \quad \text{for } \{m,n\}=\{i,j\} \text{ or } \{m,n\}=\{j,i\}. \tag{43}$$

This can be done using Lemma 4 together with an extension theorem for H^1 functions on Lipschitz domains. We note that numerical experiments suggest that no assumptions on subdomain material properties are needed, other than them being constant in each subdomain, for $\hat{C}$ in (41) to be uniformly bounded.

Our main result follows from the estimate in (42).

Theorem 1 (Condition Number Estimate). *Under the assumption in (43), the condition number of the BDDC preconditioned operator for this study is bounded by*

$$\kappa \leq C\omega^2, \tag{44}$$

where

$$\omega = \max_i(1+\log(H_i/h_i)). \tag{45}$$

In summary, we have obtained a favorable condition number estimate with less restrictive assumptions on the material properties of the subdomains than in previous studies. Comparing the condition number estimate of Theorem 1 with those in (1) and (2), we see that the factor of $H_i^2\beta_i/\alpha_i$ can be removed provided the assumption in (43) holds. In addition, the logarithmic factor has been reduced from four powers to two. We note that the estimate in Theorem 1 also holds for FETI-DP due its spectral equivalence with BDDC.

We note that the algorithm involves a non-standard averaging given by (35). This averaging requires the solution of Dirichlet problems over the union of each pair of subdomains sharing a face. The importance of this method of averaging for some problems is shown in the next section.

4 Numerical Results

In this section, we present some numerical results to verify the theory and also to provide some additional insights. The domain is a unit cube discretized into smaller

cubic elements. All the examples are solved to a relative residual tolerance of 10^{-8} for random right-hand-sides using the conjugate gradient algorithm with BDDC as the preconditioner. The number of iterations and condition number estimates from conjugate gradients are under the headings of *iter* and *cond* in the tables. We consider three different types of weights for the averaging operator. The first one, designated *SC*, is the one based on (35). Unless otherwise specified in the tables, this is the weighting used. The second type, *stiff*, is based on a conventional approach in which the weights are proportional to the entries on the diagonals of subdomain matrices. The third, *card*, uses the inverse of the cardinality of an edge, i.e. the reciprocal of the number of subdomains sharing the edge, for the weight.

The results in Table 1 are consistent with theory, suggesting condition numbers that are bounded independently of the number of subdomains, while the results in Table 2 are consistent with the $\log(H/h)^2$ estimate of Theorem 1.

We also consider a checkerboard distribution of material properties in which (α,β) for a subdomain is either (α_1,β_1) or (α_2,β_2), and note that subdomains with the same properties only share a subdomain vertex and no degrees of freedom. Results for 64 cubic subdomains each with $H/h = 4$ are shown in Table 3. Notice that for only one choice of material properties in the table do all three types of weighting lead to small condition numbers, and only the *SC* approach always gives condition numbers which are independent of the material properties. We have also investigated another type of weighting similar to *card*, but with weights γ, $0 < \gamma < 1$ for faces of subdomains with properties α_1,β_1 and $1-\gamma$ for faces of subdomains with properties α_2,β_2. Regardless of the choice of γ, large condition numbers were observed for the coefficients of the final row of Table 3. We note also that the choice of material properties in the final row is not covered by the theory of [12].

In the final example, we consider a cubic mesh of 20^3 elements that is partitioned into different numbers of subdomains using the graph partitioner Metis [8]. Although this example is not covered by our theory because the subdomains have irregular shapes, the results in Table 4 indicate that the algorithm of this study continues to perform well. The results in Tables 3 and 4 suggest that the *SC* weighting of this study may be necessary in order to effectively solve problems with material property jumps or with subdomains of irregular shape.

Table 1. Results for N cubic subdomains, each with $\beta = 1$ and $H/h = 4$.

N	$\alpha = 10^2$	$\alpha = 1$	$\alpha = 10^{-2}$
	iter (cond)	iter (cond)	iter (cond)
4^3	15 (2.70)	14 (2.63)	10 (1.77)
6^3	16 (2.88)	15 (2.81)	11 (2.05)
8^3	16 (2.95)	15 (2.87)	12 (2.23)
10^3	17 (2.98)	16 (2.91)	13 (2.33)

Table 2. Results for 64 cubic subdomains, each with $\beta = 1$.

H/h	$\alpha = 10^2$ iter (cond)	$\alpha = 1$ iter (cond)	$\alpha = 10^{-2}$ iter (cond)
4	15 (2.70)	14 (2.63)	10 (1.77)
6	17 (3.30)	16 (3.21)	11 (2.14)
8	18 (3.77)	16 (3.66)	13 (2.46)
10	19 (4.16)	18 (4.03)	13 (2.72)

Table 3. Checkerboard material property results for 64 cubic subdomains with $H/h = 4$.

α_1	β_1	α_2	β_2	*SC* iter (cond)	*stiff* iter (cond)	*card* iter (cond)
1	1	10^3	1	10 (1.59)	19 (4.57)	196 (1.64e3)
1	1	1	10^3	11 (1.96)	84 (2.69e2)	109 (4.72e2)
1	1	1	1.01	14 (2.63)	14 (2.63)	14 (2.63)
10^2	10^{-2}	1	1	6 (1.07)	65 (3.17e2)	74 (1.65e2)

Table 4. Results for 20^3 elements partitioned into N subdomains using a graph partitioner. Material properties are constant with $\alpha = 1$ and $\beta = 1$.

N	*SC* iter (cond)	*stiff* iter (cond)	*card* iter (cond)
60	19 (4.30)	189 (6.31e2)	24 (9.06)
65	19 (4.40)	184 (6.34e2)	29 (1.55e3)
70	18 (3.89)	188 (6.47e2)	23 (7.48)
75	19 (4.16)	176 (6.12e2)	23 (6.49)

Bibliography

[1] Susanne C. Brenner and Ridgway Scott. *The Mathematical Theory of Finite Element Methods*. Springer-Verlag, Berlin, Heidelberg, New York, 2008. Third edition.

[2] Clark R. Dohrmann. A preconditioner for substructuring based on constrained energy minimization. *SIAM J. Sci. Comput.*, 25(1):246–258, 2003.

[3] Clark R. Dohrmann and Olof B. Widlund. An iterative substructuring algorithm for two-dimensional problems in H(curl). *SIAM J. Numer. Anal.*, 50(3):1004–1028, 2012.

[4] Ralf Hiptmair and Jinchao Xu. Nodal auxiliary space preconditioning in H(curl) and H(div) spaces. *SIAM J. Numer. Anal.*, 45(6):2483–2509 (electronic), 2007.

[5] Ralf Hiptmair, Gisela Widmer, and Jun Zou. Auxiliary space preconditioning in $H_0(\mathrm{curl};\Omega)$. *Numer. Math.*, 103(3):435–459, 2006.

[6] Qiya Hu and Jun Zou. A nonoverlapping domain decomposition method for Maxwell's equations in three dimensions. *SIAM J. Numer. Anal.*, 41(5):1682–1708, 2003.

[7] Qiya Hu and Jun Zou. Substructuring preconditioners for saddle-point problems arising from maxwell's equations in three dimensions. *Math. Comput.*, 73 (245):35–61, 2004.

[8] George Karypis and Vipin Kumar. *METIS Version 4.0*. University of Minnesota, Department of Computer Science, Minneapolis, MN, 1998.

[9] Jing Li and Olof B. Widlund. FETI–DP, BDDC, and Block Cholesky Methods. *Internat. J. Numer. Methods Engrg.*, 66(2):250–271, 2006.

[10] Jan Mandel and Clark R. Dohrmann. Convergence of a balancing domain decomposition by constraints and energy minimization. *Numer. Linear Algebra Appl.*, 10(7):639–659, 2003.

[11] Jan Mandel, Clark R. Dohrmann, and Radek Tezaur. An algebraic theory for primal and dual substructuring methods by constraints. *Appl. Numer. Math.*, 54:167–193, 2005.

[12] Andrea Toselli. Dual–primal FETI algorithms for edge finite–element approximations in 3D. *IMA J. Numer. Anal.*, 26:96–130, 2006.

[13] Andrea Toselli and Olof Widlund. *Domain Decomposition Methods - Algorithms and Theory*, volume 34 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin Heidelberg New York, 2005.

Symbolic Techniques for Domain Decomposition Methods

T. Cluzeau[1], V. Dolean[2], F. Nataf[3], and A. Quadrat[4]

[1] Université de Limoges ; CNRS ; XLIM UMR 6172, 123 avenue Albert Thomas, 87060 Limoges cedex, France. cluzeau@ensil.unilim.fr
[2] Laboratoire J.A. Dieudonné, CNRS UMR 6621, Parc Valrose, 06108 Nice Cedex 02, France. dolean@unice.fr
[3] Laboratoire J.L. Lions, CNRS UMR 7598, Université Pierre et Marie Curie, 4 place Jussieu, 75005 Paris, France. nataf@ann.jussieu.fr
[4] INRIA Saclay-Île-de-France, DISCO Project, Supélec, L2S, 3 rue Joliot Curie, 91192 Gif-sur-Yvette cedex, France. Alban.Quadrat@inria.fr.
This work was supported by the PEPS Maths-ST2I SADDLES.

1 Introduction

Some algorithmic aspects of systems of PDEs based simulations can be better clarified by means of symbolic computation techniques. This is very important since numerical simulations heavily rely on solving systems of PDEs. For the large-scale problems we deal with in today's standard applications, it is necessary to rely on iterative Krylov methods that are scalable (i.e., weakly dependent on the number of degrees on freedom and number of subdomains) and have limited memory requirements. They are preconditioned by domain decomposition methods, incomplete factorizations and multigrid preconditioners. These techniques are well understood and efficient for scalar symmetric equations (e.g., Laplacian, biLaplacian) and to some extent for non-symmetric equations (e.g., convection-diffusion). But they have poor performances and lack robustness when used for symmetric systems of PDEs, and even more so for non-symmetric complex systems (fluid mechanics, porous media...). As a general rule, the study of iterative solvers for systems of PDEs as opposed to scalar PDEs is an underdeveloped subject.

We aim at building new robust and efficient solvers, such as domain decomposition methods and preconditioners for some linear and well-known systems of PDEs. In particular, we shall concentrate on Neumann-Neumann and FETI type algorithms which are very popular for scalar symmetric positive definite second order problems (see, for instance, [9, 11]), and to some extent to different other problems, like the advection-diffusion equations [1], plate and shell problems [16] or the Stokes equations [13]. This work is motivated by the fact that, in some sense, these methods applied to systems of PDEs (such as Stokes, Oseen, linear elasticity) are less optimal than the domain decomposition methods for scalar problems. Indeed, in the

R. Bank et al. (eds.), *Domain Decomposition Methods in Science and Engineering XX*, Lecture Notes in Computational Science and Engineering 91,
DOI 10.1007/978-3-642-35275-1_3,

case of two subdomains consisting of the two half planes, it is well-known that the Neumann-Neumann preconditioner is an exact preconditioner (the preconditioned operator is the identity operator) for the Schur complement equation for scalar equations like the Laplace problem. Unfortunately, this does not hold in the vector case.

In order to achieve this goal, we use algebraic methods developed in constructive algebra, D-modules (differential modules) and symbolic computation such as the so-called Smith or Jacobson normal forms and Gröbner basis techniques for transforming a linear system of PDEs into a set of independent PDEs. These algebraic and symbolic methods provide important intrinsic information (e.g., invariants) about the linear system of PDEs to solve. These build-in properties need to be taken into account in the design of new numerical methods, which can supersede the usual ones based on a direct extension of the classical scalar methods to linear systems of PDEs.

By means of these techniques, it is also possible to transform the linear system of PDEs into a set of decoupled PDEs under certain types of invertible transformations. One of these techniques is the so-called Smith normal form of the matrix of OD operators associated with the linear system. This normal form was introduced by H. J. S. Smith (1826–1883) for matrices with integer entries (see, e.g., [17], Theorem 1.4). The Smith normal form has already been successfully applied to open problems in the design of Perfectly Matched Layers (PML). The theory of PML for scalar equations was well-developed and the usage of the Smith normal form allowed to extend these works to systems of PDEs. In [12], a general approach is proposed and applied to the particular case of the compressible Euler equations that model aeroacoustic phenomena and in [2] for shallow-water equations.

For domain decomposition methods, several results have been obtained on compressible Euler equations [7], Stokes and Oseen systems [8] or in [10] where a new method in the "Smith" spirit has been derived. Previously the computations were performed heuristically, whereas in this work, we aim at finding a systematic way to build optimal algorithms for given PDE systems.

Notations. If R is a ring, then $R^{p\times q}$ is the set of $p\times q$ matrices with entries in R and $\mathrm{GL}_p(R)$ is the group of invertible matrices of $R^{p\times p}$, namely $\mathrm{GL}_p(R)=\{E\in R^{p\times p}\mid \exists\, F\in R^{p\times p}: EF=FE=I_p\}$. An element of $\mathrm{GL}_p(R)$ is called an *unimodular matrix*. A diagonal matrix with elements d_i's will be denoted by $\mathrm{diag}(d_1,\ldots,d_p)$. If k is a field (e.g., $k=\mathbb{Q},\mathbb{R},\mathbb{C}$), then $k[x_1,\ldots,x_n]$ is the commutative ring of polynomials in $x_1,\ldots,x_n$ with coefficients in k. In what follows, $k(x_1,\ldots,x_n)$ will denote the field of rational functions in $x_1,\ldots,x_n$ with coefficients in k. Finally, if $r, r'\in R$, then $r'\,|\,r$ means that r' divides r, i.e., there exists $r''\in R$ such that $r=r''r'$.

2 Smith Normal Form of Linear Systems of PDEs

We first introduce the concept of *Smith normal form* of a matrix with polynomial entries (see, e.g., [17], Theorem 1.4). The Smith normal form is a mathematical technique which is classically used in module theory, linear algebra, symbolic computation, ordinary differential systems, and control theory. It was first developed to study matrices with integer entries. But, it was proved to exist for any *principal ideal*

domain (namely, a commutative ring R whose ideals can be generated by an element of R) [15]. Since $R = k[s]$ is a principal ideal domain when k is a field, we have the following theorem only stated for square matrices.

Theorem 1. *Let k be a field, $R = k[s]$, p a positive integer and $A \in R^{p\times p}$. Then, there exist two matrices $E \in \mathrm{GL}_p(R)$ and $F \in \mathrm{GL}_p(R)$ such that*

$$A = E\,S\,F,$$

where $S = \mathrm{diag}(d_1, \ldots, d_p)$ and the $d_i \in R$ satisfying $d_1 \,|\, d_2 \,|\, \cdots \,|\, d_p$. In particular, we can take $d_i = m_i/m_{i-1}$, where m_i is the greatest common divisor of all the $i \times i$-minors of A (i.e., the determinants of all $i \times i$-submatrices of A), with the convention that $m_0 = 1$. The matrix $S = \mathrm{diag}(d_1, \ldots, d_p) \in R^{p\times p}$ is called a Smith normal form *of A.*

We note that $E \in \mathrm{GL}_p(R)$ is equivalent to $\det(E)$ is an invertible polynomial, i.e., $\det(E) \in k \setminus \{0\}$. Also, in what follows, we shall assume that the d_i's are *monic polynomials*, i.e., their leading coefficients are 1, which will allow us to call the matrix $S = \mathrm{diag}(d_1, \ldots, d_p)$ *the* Smith normal form of A. But, the unimodular matrices E and F are not uniquely defined by A. The proof of Theorem 1 is constructive and gives an algorithm for computing matrices E, S and F. The computation of Smith normal forms is available in many computer algebra systems such as `Maple`, `Mathematica`, `Magma`...

Consider now the following model problem in $\mathbb{R}^d$ with $d = 2, 3$:

$$\mathscr{L}_d(\mathbf{w}) = \mathbf{g} \quad \text{in } \mathbb{R}^d, \quad |\mathbf{w}(\mathbf{x})| \to 0 \quad \text{for } |\mathbf{x}| \to \infty. \tag{1}$$

For instance, $\mathscr{L}_d(\mathbf{w})$ can represent the Stokes/Oseen/linear elasticity operators in dimension d. Moreover, if we suppose that the inhomogeneous linear system of PDEs (1) has constant coefficients, then it can be rewritten as

$$A_d\,\mathbf{w} = \mathbf{g}, \tag{2}$$

where $A_d \in R^{p\times p}$, $R = k[\partial_x, \partial_y]$ (resp., $R = k[\partial_x, \partial_y, \partial_z]$) for $d = 2$ (resp., $d = 3$) and k is a field.

In what follows, we shall study the domain decomposition problem in which $\mathbb{R}^d$ is divided into subdomains. We assume that the direction normal to the interface of the subdomains is particularized and denoted by ∂_x. If $R_x = k(\partial_y)[\partial_x]$ for $d = 2$ or $R_x = k(\partial_y, \partial_z)[\partial_x]$ for $d = 3$, then, computing the Smith normal form of the matrix $A_d \in R_x^{p\times p}$, we obtain $A_d = E\,S\,F$, where $S \in R_x^{p\times p}$ is a diagonal matrix, $E \in \mathrm{GL}_p(R_x)$ and $F \in \mathrm{GL}_p(R_x)$. The entries of the matrices E, S, F are polynomials in ∂_x, and E and F are unimodular matrices, i.e., $\det(E), \det(F) \in k(\partial_y) \setminus \{0\}$ if $d = 2$, or $\det(E), \det(F) \in k(\partial_y, \partial_z) \setminus \{0\}$ if $d = 3$. We recall that the matrices E and F are not unique contrary to S. Using the Smith normal form of A_d, we get:

$$A_d\,\mathbf{w} = \mathbf{g} \quad \Leftrightarrow \quad \{\mathbf{w_s} := F\,\mathbf{w},\ S\,\mathbf{w_s} = E^{-1}\,\mathbf{g}\}. \tag{3}$$

In other words, (3) is equivalent to the uncoupled linear system:

$$S\,\mathbf{w_s} = E^{-1}\,\mathbf{g}. \tag{4}$$

Since $E \in \mathrm{GL}_p(R_x)$ and $F \in \mathrm{GL}_p(R_x)$, the entries of their inverses are still polynomial in ∂_x. Thus, applying E^{-1} to the right-hand side $\mathbf{g}$ of $A_d\,\mathbf{w} = \mathbf{g}$ amounts to taking k-linear combinations of derivatives of $\mathbf{g}$ with respect to x. If $\mathbb{R}^d$ is split into two subdomains $\mathbb{R}^- \times \mathbb{R}^{d-1}$ and $\mathbb{R}^+ \times \mathbb{R}^{d-1}$, where $\mathbb{R}^- = \{x \in \mathbb{R} \mid x < 0\}$ and $\mathbb{R}^+ = \{x \in \mathbb{R} \mid x > 0\}$, then the application of E^{-1} and F^{-1} to a vector can be done for each subdomain independently. No communication between the subdomains is necessary.

In conclusion, it is enough to find a domain decomposition algorithm for the uncoupled system (4) and then transform it back to the original one (2) by means of the invertible matrix F over R_x. This technique can be applied to any linear system of PDEs once it is rewritten in a polynomial form. The uncoupled system acts on the new dependent variables $\mathbf{w_s}$, which we shall further call *Smith variables* since they are issued from the Smith normal form.

Remark 1. Since the matrix F is used to transform (4) to (2) (see the first equation of the right-hand side of (3)) and F is not unique, we need to find a matrix F as simple as possible (e.g., F has minimal degree in ∂_x) so that to obtain a final algorithm whose form can be used for practical computations.

Example 1 Consider the two dimensional elasticity operator defined by $\mathscr{E}_2(\mathbf{u}) := -\mu\,\Delta\,\mathbf{u} - (\lambda+\mu)\,\nabla\,\mathrm{div}\,\mathbf{u}$. If we consider the commutative polynomial rings $R = \mathbb{Q}(\lambda,\mu)[\partial_x,\partial_y]$, $R_x = \mathbb{Q}(\lambda,\mu)(\partial_y)[\partial_x] = \mathbb{Q}(\lambda,\mu,\partial_y)[\partial_x]$ and

$$A_2 = \begin{pmatrix} (\lambda+2\,\mu)\,\partial_x^2 + \mu\,\partial_y^2 & (\lambda+\mu)\,\partial_x\,\partial_y \\ (\lambda+\mu)\,\partial_x\,\partial_y & \mu\,\partial_x^2 + (\lambda+2\,\mu)\,\partial_y^2 \end{pmatrix} \in R^{2\times 2}$$

the matrix of PD operators associated with $\mathscr{E}_2$, i.e., $\mathscr{E}_2(\mathbf{u}) = A_2\,\mathbf{u}$, then the Smith normal form of $A_2 \in R_x^{2\times 2}$ is defined by:

$$S_{A_2} = \begin{pmatrix} 1 & 0 \\ 0 & \Delta^2 \end{pmatrix}. \tag{5}$$

The particular form of S_{A_2} shows that, over R_x, the system of PDEs for the linear elasticity in $\mathbb{R}^2$ is algebraically equivalent to a biharmonic equation.

Example 2 Consider the two dimensional Oseen operator $\mathscr{O}_2(\mathbf{w}) = \mathscr{O}_2(\mathbf{v},q) := (c\,\mathbf{v} - \nu\,\Delta\mathbf{v} + \mathbf{b}\cdot\nabla\mathbf{v} + \nabla q, \nabla\cdot\mathbf{v})$, where $\mathbf{b}$ is the convection velocity. If $\mathbf{b} = 0$, then we obtain the Stokes operator $\mathscr{S}_2(\mathbf{w}) = \mathscr{S}_2(\mathbf{v},q) := (c\,\mathbf{v} - \nu\,\Delta\mathbf{v} + \nabla q, \nabla\cdot\mathbf{v})$. If $R = \mathbb{Q}(b_1,b_2,c,\nu)[\partial_x,\partial_y]$, $R_x = \mathbb{Q}(b_1,b_2,c,\nu)(\partial_y)[\partial_x] = \mathbb{Q}(b_1,b_2,c,\nu,\partial_y)[\partial_x]$ and

$$O_2 = \begin{pmatrix} -\nu\,(\partial_x^2+\partial_y^2) + b_1\,\partial_x + b_2\,\partial_y + c & 0 & \partial_x \\ 0 & -\nu\,(\partial_x^2+\partial_y^2) + b_1\,\partial_x + b_2\,\partial_y + c & \partial_y \\ \partial_x & \partial_y & 0 \end{pmatrix}$$

the matrix of PD operators associated with $\mathscr{O}_2$, i.e., $\mathscr{O}_2(\mathbf{w}) = O_2\,\mathbf{w}$, then the Smith normal form of $O_2 \in R_x^{3\times 3}$ is defined by:

$$S_{O_2} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \Delta L_2 \end{pmatrix}, \quad L_2 = c - \nu\Delta + \mathbf{b}\cdot\nabla. \tag{6}$$

From the form of S_{O_2} we can deduce that the two-dimensional Oseen equations can be mainly characterized by the scalar fourth order PD operator ΔL_2. This is not surprising since the stream function formulation of the Oseen equations for $d = 2$ gives the same PDE for the stream function.

Remark 2. The above applications of Smith normal forms suggest that one should design an optimal domain decomposition method for the biharmonic operator Δ^2 (resp., $L_2\Delta$) in the case of linear elasticity (resp., the Oseen/Stokes equations) for the two-dimensional problems, and then transform it back to the original system.

3 An Optimal Algorithm for the Biharmonic Operator

We give here an example of Neumann-Neumann methods in its iterative version for Laplace and biLaplace equations. For simplicity, consider a decomposition of the domain $\Omega = \mathbb{R}^2$ into two half planes $\Omega_1 = \mathbb{R}^- \times \mathbb{R}$ and $\Omega_2 = \mathbb{R}^+ \times \mathbb{R}$. Let the interface $\{0\} \times \mathbb{R}$ be denoted by Γ and $(\mathbf{n}_i)_{i=1,2}$ be the outward normal of $(\Omega_i)_{i=1,2}$. We consider the following problem:

$$-\Delta u = f \text{ in } \mathbb{R}^2, \quad |u(\mathbf{x})| \to 0 \text{ for } |\mathbf{x}| \to \infty. \tag{7}$$

and the following **Neumann-Neumann algorithm** applied to problem (7):
Let u_Γ^n be the interface solution at iteration n. We obtain u_Γ^{n+1} from u_Γ^n by the following iterative procedure

$$\begin{cases} -\Delta u^{i,n} = f, & \text{in } \Omega_i, \\ u^{i,n} = u_\Gamma^n, & \text{on } \Gamma, \end{cases} \quad \begin{cases} -\Delta \tilde{u}^{i,n} = 0, & \text{in } \Omega_i, \\ \dfrac{\partial \tilde{u}^{i,n}}{\partial \mathbf{n}_i} = -\dfrac{1}{2}\left(\dfrac{\partial u^{1,n}}{\partial \mathbf{n}_1} + \dfrac{\partial u^{2,n}}{\partial \mathbf{n}_2}\right), & \text{on } \Gamma, \end{cases} \tag{8}$$

and then $u_\Gamma^{n+1} = u_\Gamma^n + \frac{1}{2}\left(\tilde{u}^{1,n} + \tilde{u}^{2,n}\right)$.
This algorithm is *optimal* in the sense that it converges in two iterations.

Since the biharmonic operator seems to play a key role in the design of a new algorithm for both Stokes and elasticity problem in two dimensions, we need to build an optimal algorithm for it. We consider the following problem:
Find $\phi : \mathbb{R}^2 \to \mathbb{R}$ such that:

$$\Delta^2\phi = g \text{ in } \mathbb{R}^2, \quad |\phi(\mathbf{x})| \to 0 \text{ for } |\mathbf{x}| \to \infty. \tag{9}$$

and the following **"Neumann-Neumann" type algorithm** applied to (9):
Let $(\phi_\Gamma^n, D\phi_\Gamma^n)$ be the interface solution at iteration n (suppose also that $\phi_\Gamma^0 =$

$\phi^0|_\Gamma, D\phi^0_\Gamma = (\Delta\phi^0)_\Gamma$). We obtain $(\phi^{n+1}_\Gamma, D\phi^n_\Gamma)$ from $(\phi^n_\Gamma, D\phi^n_\Gamma)$ by the following iterative procedure

$$\left\{\begin{array}{lll} -\Delta^2\phi^{i,n} = f, & & \text{in } \Omega_i, \\ \phi^{i,n} = \phi^n_\Gamma, & & \text{on } \Gamma, \\ \Delta\phi^{i,n} = D\phi^n_\Gamma, & & \text{on } \Gamma, \end{array}\right. \quad \left\{\begin{array}{lll} -\Delta^2\tilde{\phi}^{i,n} = 0, & & \text{in } \Omega_i, \\ \dfrac{\partial\tilde{\phi}^{i,n}}{\partial\mathbf{n}_i} = -\dfrac{1}{2}\left(\dfrac{\partial\phi^{1,n}}{\partial\mathbf{n}_1} + \dfrac{\partial\phi^{2,n}}{\partial\mathbf{n}_2}\right), & & \text{on } \Gamma, \\ \dfrac{\partial\Delta\tilde{\phi}^{i,n}}{\partial\mathbf{n}_i} = -\dfrac{1}{2}\left(\dfrac{\partial\Delta\phi^{1,n}}{\partial\mathbf{n}_1} + \dfrac{\partial\Delta\phi^{2,n}}{\partial\mathbf{n}_2}\right), & & \text{on } \Gamma, \end{array}\right. \tag{10}$$

and then $\phi^{n+1}_\Gamma = \phi^n_\Gamma + \frac{1}{2}\left(\tilde{\phi}^{1,n} + \tilde{\phi}^{2,n}\right)$, $D\phi^{n+1}_\Gamma = D\phi^n_\Gamma + \frac{1}{2}\left(\tilde{\Delta}\phi^{1,n} + \tilde{\Delta}\phi^{2,n}\right)$.

This is a generalization of the Neumann-Neumann algorithm for the Δ operator and is also *optimal* (the proof can be found in [8]).

Now, in the case of the two dimensional linear elasticity, ϕ represents the second component of the vector of Smith variables, that is, $\phi = (\mathbf{w}_s)_2 = (F\mathbf{u})_2$, where $\mathbf{u} = (u, v)$ is the displacement field. Hence, we need to replace ϕ with $(F\mathbf{u})_2$ into the algorithm for the biLaplacian, and then simplify it using algebraically admissible operations. Thus, one can obtain an optimal algorithm for the Stokes equations or linear elasticity depending on the form of F. From here comes the necessity of choosing in a proper way the matrix F (which is not unique), used to define the Smith normal form, in order to obtain a "good" algorithm for the systems of PDEs from the optimal one applied to the biharmonic operator. In [7] and [8], the computation of the Smith normal forms for the Euler equations and the Stokes equations was done by hand or using the `Maple` command *Smith*. Surprisingly, the corresponding matrices F have provided good algorithms for the Euler equations and the Stokes equations even if the approach was entirely heuristic.

4 Relevant Smith Variables: A Completion Problem

The efficiency of our algorithms heavily relies on the simplicity of the Smith variables, that is on the entries of the unimodular matrix F used to compute the Smith normal form of the matrix A. In this section, within a constructive *algebraic analysis* approach, we develop a method for constructing many possible Smith variables. Taking into account physical aspects, the user can then choose the simplest one among them. We are going to show that the problem of finding Smith variables can be reduced to a *completion problem*. First of all, we very briefly introduce some notions of module theory [15].

Given a ring R (e.g., $R = k[\partial_1, \ldots, \partial_d]$, where k is a field (e.g., $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$)), the definition of a *R-module* M is similar to the one of a vector space but where the scalars are taken in the ring R and not in a field as for vector spaces. If $A \in R^{p\times p}$, then the kernel of the R-linear map (*R-homomorphism*) $.A : R^{1\times p} \longrightarrow R^{1\times p}$, defined by $(.A)(\mathbf{r}) = \mathbf{r}A$, is the R-module defined by:

$$\ker_R(.A) = \{\mathbf{r} \in R^{1\times p} \mid \mathbf{r}A = 0\}.$$

The image $\mathrm{im}_R(.A)$ of $.A$, simply denoted by $R^{1\times p}A$, is the R-module defined by all the R-linear combinations of the rows of A. The cokernel $\mathrm{coker}_R(.A)$ of $.A$ is the *factor* R-module defined by $\mathrm{coker}_R(.A) = R^{1\times p}/(R^{1\times p}A)$. To simplify the notation, we shall denote this module by M. M is nothing more than the R-module of the row vectors of $R^{1\times p}$ modulo the R-linear combinations of rows of A. Let $R_1 = k(\partial_2,\ldots,\partial_d)[\partial_1]$, $R_i = k(\partial_1,\ldots,\partial_{i-1},\partial_{i+1},\ldots,\partial_d)[\partial_i]$, $i = 2,\ldots,d-1$, and $R_d = k(\partial_1,\ldots,\partial_{d-1})[\partial_d]$ be the polynomial rings in ∂_i with coefficients in the field of rational functions in all other PD operators.

Since the R-module $M = R^{1\times p}/(R^{1\times p}A)$ plays a fundamental role in what follows, let us describe it in terms of generators and relations. Let $\{\mathbf{f_j}\}_{j=1,\ldots,p}$ be the standard basis of $R^{1\times p}$, namely $\mathbf{f_j}$ is the row vector of $R^{1\times p}$ defined by 1 at the jth position and 0 elsewhere, and m_j the residue class of $\mathbf{f_j}$ in M. Then, $\{m_j\}_{j=1,\ldots,p}$ is a family of generators of the R-module M, i.e., for any $m \in M$, then there exists $\mathbf{r} = (r_1,\ldots,r_p) \in R^{1\times p}$ such that $m = \sum_{j=1}^{p} r_j m_j$ [3]. The family of generators $\{m_j\}_{j=1,\ldots,p}$ of M satisfies the relations $\sum_{j=1}^{p} A_{ij} m_j = 0$ for all $i = 1,\ldots,p$ [3]. For more details, see [3, 15].

Let $E, F \in \mathrm{GL}_p(R_i)$ be two unimodular matrices such that $A = ESF$, where $S = \mathrm{diag}(1,\ldots,1,d_{r+1},\ldots,d_q)$ is the Smith normal form of A. Moreover, let us split $F \in \mathrm{GL}_p(R_i)$ into two parts row-wise, i.e., $F = (F_1^T \quad F_2^T)^T$, where $F_1 \in R_i^{r\times p}$, $F_2 \in R_i^{(p-r)\times p}$, and r is the number of ones in S. Then:

$$A = ESF \quad \Leftrightarrow \quad \begin{pmatrix} F_1 \\ S_2 F_2 \end{pmatrix} = E^{-1}A, \quad S_2 = \mathrm{diag}(d_{r+1},\ldots,d_p). \tag{11}$$

Cleaning the denominators of the entries of S_2 (resp., F_2), we can assume without loss of generality that the d_j's (resp., the entries of F_2) belong to R. Then, (11) shows that the jth row of F_2 must be an element of the R_i-module $M_i = R_i^{1\times p}/(R_i^{1\times p}A)$ annihilated by d_j. Consequently, the possible F_2's can be found by computing a family of generators of the R_i-modules $\mathrm{ann}_{M_i}(d_j) = \{m \in M_i \mid d_j m = 0\}$ for $j = r+1,\ldots,p$. These R_i-modules can be computed by means of *Gröbner basis techniques* (see, e.g., [6]). Hence, we get $S_2F_2 = G_2A$ for some $G_2 \in R_i^{(p-r)\times p}$. Then, for each choice for F_2, we are reduced to the following *completion problem*:

$$\begin{aligned} &\text{Find } F_1 \in R_i^{r\times p} \text{ such that } F = (F_1^T \quad F_2^T)^T \in \mathrm{GL}_p(R_i) \text{ and } F_1 = G_1 A \\ &\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{for some } G_1 \in R_i^{r\times p}. \end{aligned} \tag{12}$$

Example 3 Let $R = \mathbb{Q}(\lambda,\mu)[\partial_x,\partial_y,\partial_z]$ be the commutative polynomial ring of PD operators in ∂_x, ∂_y and ∂_z with coefficients in the field $\mathbb{Q}(\lambda,\mu)$,

$$A = \begin{pmatrix} -(\lambda+\mu)\,\partial_x^2 - \mu\,\Delta & -(\lambda+\mu)\,\partial_x\partial_y & -(\lambda+\mu)\,\partial_x\partial_z \\ -(\lambda+\mu)\,\partial_x\partial_y & -(\lambda+\mu)\,\partial_y^2 - \mu\,\Delta & -(\lambda+\mu)\,\partial_y\partial_z \\ -(\lambda+\mu)\,\partial_x\partial_z & -(\lambda+\mu)\,\partial_y\partial_z & -(\lambda+\mu)\,\partial_z^2 - \mu\,\Delta \end{pmatrix} \in R^{3\times 3}$$

the matrix of PD operators defining the elastostatic equations in $\mathbb{R}^3$, where $\Delta = \partial_x^2 + \partial_y^2 + \partial_z^2$, and the associated R-module $M = R^{1\times 3}/(R^{1\times 3}A)$. The Smith normal form

of A with respect to x is given by $S = \mathrm{diag}(1, \Delta, \Delta^2)$. With the above notations, we get $r = 1$ and $S_2 = \mathrm{diag}(\Delta, \Delta^2) \in R^{2\times 2}$. Let $R_x = \mathbb{Q}(\lambda, \mu)(\partial_y, \partial_z)[\partial_x]$, $F_1 \in R_x^{1\times 3}$ and $F_2 \in R_x^{2\times 3}$. Then, the first (resp. second) row of F_2 must be an element of the R_x-module $M_x = R_x^{1\times 3}/(R_x^{1\times 3} A)$ annihilated by $\Delta \in R$ (resp. $\Delta^2 \in R$). Using the OREMODULES package [4], we find that families of generators of $\mathrm{ann}_{M_x}(\Delta)$ and $\mathrm{ann}_{M_x}(\Delta^2)$ are respectively defined by the residue classes of the rows of the following matrices in M_x:

$$A_\Delta = \begin{pmatrix} 0 & -\partial_z & \partial_y \\ \partial_z & 0 & -\partial_x \\ -\partial_y & \partial_x & 0 \\ \partial_x & \partial_y & \partial_z \end{pmatrix}, \quad A_{\Delta^2} = I_3.$$

That simply means that a family of generators of $\mathrm{ann}_{M_x}(\Delta)$ is given by the divergence and the curl of the displacement field and for $\mathrm{ann}_{M_x}(\Delta^2)$ by the components of the displacement fields. Now, the first (resp., second) row of F_2 must be a R_x-linear combination of the rows of A_Δ (resp., A_{Δ^2}). We thus have several choices and for each of them, we are reduced to a completion problem (12). For instance, choosing the first row of A_Δ (resp., the third row of A_{Δ^2}) as first (resp., second) row of F_2, namely

$$F_2 = \begin{pmatrix} 0 & -\partial_z & \partial_y \\ 0 & 0 & 1 \end{pmatrix},$$

we then have to find a row vector $F_1 \in R_x^{1\times 3}$ such that $F_1 = G_1 A$ for some $G_1 \in R_x^{1\times 3}$ and $F = (F_1^T \quad F_2^T)^T \in \mathrm{GL}_3(R_x)$. If such a row vector F_1 exists, then the matrix $F = (F_1^T \quad F_2^T)^T$ provides a good choice of Smith variables.

We first give two necessary conditions for a choice of F_2 to provide a solution of the completion problem (straightforward from the relation $A = ESF$):

Lemma 1. *With the above notations, given $F_2 \in R^{(p-r)\times p}$, necessary conditions for the solvability of the completion problem (12) are:*

1. *F_2 admits a right inverse over R_i, i.e. $\exists\, S_2 \in R_i^{p\times(p-r)} :\ F_2 S_2 = I_{p-r}$.*
2. *There exists a matrix $G_2 \in R_i^{(p-r)\times p}$ such that $S_2 F_2 = G_2 A$.*

Since R_i is a *principal ideal domain* (namely, every ideal of R_i can be generated by an element of R_i), Condition 1 of Lemma 1 is equivalent to the condition that the R_i-module $\mathrm{coker}_{R_i}(.F_2) = R_i^{1\times p}/(R_i^{1\times(p-r)} F_2)$ is free of rank r, i.e. $\mathrm{coker}_{R_i}(.F_2)$ admits a basis of cardinality r [3, 15]. It is equivalent to the existence of two matrices $Q_2 \in R_i^{p\times r}$ and $T_2 \in R_i^{r\times p}$ such that $\ker_{R_i}(.Q_2) = R_i^{1\times(p-r)} F_2$ and $T_2 Q_2 = I_r$ [3]. Such a matrix Q_2 is called an *injective parametrization* of $\mathrm{coker}_{R_i}(.F_2)$. Matrices Q_2 and T_2 can be computed by Gröbner basis techniques [3]. The corresponding algorithms are implemented in the OREMODULES package [4]. The next theorem characterizes the solvability of the completion problem (12).

Theorem 2. *Let $F_2 \in R^{(p-r)\times p}$ admit a right inverse over R_i and satisfy $S_2 F_2 = G_2 A$ for some $G_2 \in R_i^{(p-r)\times p}$. If Q_2 is an injective parametrization of the free R_i-module $\mathrm{coker}_{R_i}(.F_2)$ of rank r, and $T_2 \in R_i^{r\times p}$ a left inverse of Q_2, then a necessary and sufficient condition for the existence of a solution of the completion problem (12) is the existence of two matrices $H \in R_i^{r\times(p-r)}$ and $G_1 \in R_i^{r\times p}$ such that $T_2 = G_1 A - H F_2$. Then, $F_1 = T_2 + H F_2 = G_1 A$ is a solution of the completion problem (12), i.e., $F = ((T_2 + H F_2)^T \quad F_2^T)^T \in \mathrm{GL}_p(R_i)$ is such that $A = E S F$ for some $E \in \mathrm{GL}_p(R_i)$, where S is the Smith normal form of A.*

From the explanations above, we deduce the following algorithm that, given A, $S_2 = \mathrm{diag}(d_{r+1}, \ldots, d_p)$, and a choice for F_2 computed from the calculations of $\mathrm{ann}_{M_i}(d_j)$ for $d_j \in R$, find (if it exists) a completion of F_2. The following algorithm

---

Input: $A \in R^{p\times p}$, $S_2 \in R^{(p-r)\times(p-r)}$ and $F_2 \in R^{(p-r)\times p}$.
Output: A completion $F = (F_1^T \quad F_2^T)^T$ of F_2 or "No completion exists".

1. Compute a right inverse of F_2 over R_i;
2. **If** no right inverse exists, then RETURN "No completion exists", **Else**
   (a) Factorize $S_2 F_2$ with respect to A over R_i;
   (b) **If** no factorization exists, then RETURN "No completion exists", **Else**
      i. Compute an injective parametrization Q_2 of $\mathrm{coker}_{R_i}(.F_2)$;
      ii. Compute a left inverse T_2 of Q_2 over R_i;
      iii. Factorize T_2 with respect to $(F_2^T \quad A^T)^T$ over R_i;
      iv. **If** no factorization exists, then RETURN "No completion exists", **Else**
      note $T_2 = (-H \quad G_1)\begin{pmatrix} F_2 \\ A \end{pmatrix}$ and RETURN $F = \begin{pmatrix} T_2 + H F_2 \\ F_2 \end{pmatrix}$.

---

was implemented in `Maple` based on the OREMODULES package.

Example 4 Consider again the elastostatic equations introduced in Example 3. For the choice of F_2 given at the end of Example 3, our implementation succeeds in finding a completion and we get the following completion of F_2:

$$F = \begin{pmatrix} 1 & -\frac{\partial_x \partial_y}{\partial_y^2 + \partial_z^2} & -\frac{\partial_x((\lambda+2\mu)(\partial_x^2+\partial_y^2)+(2\lambda+3\mu)\partial_z^2)}{(\lambda+\mu)\partial_z(\partial_y^2+\partial_z^2)} \\ 0 & -\partial_z & \partial_y \\ 0 & 0 & 1 \end{pmatrix} \in \mathrm{GL}_3(R_x).$$

For more details and explicit computations, we refer the reader to [5].

5 Reduction of the Interface Conditions

In the algorithms presented in the previous sections, we have equations in the domains Ω_i and interface conditions on Γ obtained heuristically. We need to find an

automatic way to reduce the interface conditions with respect to the equations in the domains. In this section, we show how symbolic computations can be used to perform such reductions. The naïve idea consists in gathering all equations and compute a Gröbner basis [6]. However, one has to keep in mind that the independent variables do not play the same role. More precisely, the interface conditions cannot be differentiated with respect to x since the border of the interface is defined by $x = 0$. Consequently, we have developed and implemented an alternative method in `Maple` using the OREMODULES package, which can be sketched as follows:

1. Compute a Gröbner basis of the polynomial equations inside the domain for a relevant monomial order;
2. Compute the normal forms of the interface conditions with respect to the latter Gröbner basis;
3. Write these normal forms in the *jet notations* with respect to the independent variable x, i.e., rewrite the derivatives $\partial_x^i y_k$ of the dependent variables y_k as new indeterminates $y_{k,i}$;
4. Perform linear algebra manipulations to simplify the normal forms.

For more details and explicit computations, we refer the reader to [5].

6 Some Optimal Algorithms

After performing the completion and the reduction of the interface conditions, we can give examples of optimal algorithms (elasticity and Stokes equations).

Example 5 Consider the elasticity operator:

$$\mathscr{E}_d\,\mathbf{u} = -\operatorname{div}\sigma(\mathbf{u}), \quad \sigma(\mathbf{u}) = \mu\,(\nabla\mathbf{u} + (\nabla\mathbf{u})^T) + \lambda\,\operatorname{div}\mathbf{u}\,I_d.$$

If $d = 2$, then the completion algorithm gives two possible choices for F:

$$F = \begin{pmatrix} -\frac{\partial_x\,(\mu\,\partial_x^2 - \lambda\,\partial_y^2)}{(\lambda+\mu)\,\partial_y^3} & 1 \\ 1 & 0 \end{pmatrix}, \quad F = \begin{pmatrix} 1 & -\frac{(\lambda+\mu)\partial_x\,((3\,\mu+2\,\lambda)\,\partial_y^2 + (2\,\mu+\lambda)\,\partial_x^2)}{\partial_y^3} \\ 0 & 1 \end{pmatrix}. \tag{13}$$

By replacing ϕ into the Neumann-Neumann algorithm for the biLaplacian by $(F\mathbf{u})_2$ and re-writing the interface conditions, using the equations inside the domain like in [8], we get two different algorithms for the elasticity system. Note that, in the first case of (13), $\phi = u$, and, in the second one, $\phi = v$ (where $\mathbf{u} = (u, v)$). Below, we shall write in detail the algorithm in the second case. To simplify the writing, we denote by $u_\tau = \mathbf{u}\cdot\tau$, $u_\mathbf{n} = \mathbf{u}\cdot\mathbf{n}$, $\sigma_{\mathbf{nn}}(\mathbf{u}) = (\sigma(\mathbf{u})\cdot\mathbf{n})\cdot\mathbf{n}$, $\sigma_{\mathbf{n}\tau}(\mathbf{u}) = (\sigma(\mathbf{u})\cdot\mathbf{n})\cdot\tau$.

Let $(u_\Gamma^n, \sigma_\Gamma^n)$ be the interface solution at iteration n (suppose also that $u_\Gamma^0 = (u_\tau^0)|_\Gamma$, $\sigma_\Gamma^0 = (\sigma_{s\mathbf{nn}}(u^0))|_\Gamma$). We obtain $(u_\Gamma^{n+1}, \sigma_\Gamma^n)$ from $(u_\Gamma^n, \sigma_\Gamma^n)$ by the following iterative procedure

$$\begin{cases} \mathscr{E}_2(\mathbf{u}^{i,n}) &= f, \quad in\ \Omega_i, \\ u^{1,n}_{\tau_i} &= u^n_\Gamma, \ on\ \Gamma, \\ \sigma_{\mathbf{n}_i\mathbf{n}_i}(\mathbf{u}^{i,n}) &= \sigma^n_\Gamma, \ on\ \Gamma, \end{cases} \begin{cases} \mathscr{E}_2(\tilde{\mathbf{u}}^{i,n}) &= 0, & in\ \Omega_i, \\ \tilde{\mathbf{u}}^{i,n}_{\tau_i} &= -\frac{1}{2}\left(\mathbf{u}^{1,n}_{\mathbf{n}_1} + \mathbf{u}^{2,n}_{\mathbf{n}_2}\right), & on\ \Gamma, \\ \sigma_{\mathbf{n}_i\tau_i}(\tilde{\mathbf{u}}^{i,n}) &= -\frac{1}{2}\left(\sigma_{\mathbf{n}_1\tau_1}(\mathbf{u}^{1,n}) + \sigma_{\mathbf{n}_2\tau_2}(\mathbf{u}^{2,n})\right), & \\ & & on\ \Gamma, \end{cases} \tag{14}$$

and $u^{n+1}_\Gamma = u^n_\Gamma + \frac{1}{2}\left(\tilde{u}^{1,n}_{\tau_1} + \tilde{u}^{2,n}_{\tau_2}\right)$, $\sigma^{n+1}_\Gamma = \sigma^n_\Gamma + \frac{1}{2}\left(\sigma_{\mathbf{n}_1\mathbf{n}_1}(\tilde{\mathbf{u}}^{1,n}) + \sigma_{\mathbf{n}_2\mathbf{n}_2}(\tilde{\mathbf{u}}^{2,n})\right)$.

Remark 3. We found an algorithm with a mechanical meaning: Find the tangential part of the normal stress and the normal displacement at the interface so that the normal part of the normal stress and the tangential displacement on the interface match. This is very similar to the original Neumann-Neumann algorithm, which means that the implementation effort of the new algorithm from an existing Neumann-Neumann is negligible (the same type of quantities – displacement fields and efforts – are imposed at the interfaces), except that the new algorithm requires the knowledge of some geometric quantities, such as normal and tangential vectors. Note also that, with the adjustment of the definition of tangential quantities for $d = 3$, the algorithm is the same, and is also similar to the results in [8].

7 Conclusion

All algorithms and interface conditions are derived for problems posed on the whole space, since for the time being, this is the only way to treat from the algebraic point of view these problems. The effect of the boundary condition on bounded domains cannot be quantified with the same tools. All the algorithms are designed in the PDE level and it is very important to choose the right discrete framework in order to preserve the optimal properties. For example, in the case of linear elasticity a good candidate would be the TDNNS finite elements that can be found in [14]. The implementation and the impact of the discretizations on the algorithms is an ongoing work.

Bibliography

[1] Y. Achdou, P. Le Tallec, F. Nataf, and M. Vidrascu. A domain decomposition preconditioner for an advection-diffusion problem. *Comput. Methods Appl. Mech. Engrg*, 184:145–170, 2000.

[2] H. Barucq, J. Diaz, and M. Tlemcani. New absorbing layers conditions for short water waves. *Journal of Computational Physics*, 229(1):58–72, 2010.

[3] F. Chyzak, A. Quadrat, and D. Robertz. Effective algorithms for parametrizing linear control systems over Ore algebras. *Appl. Algebra Engrg. Comm. Comput.*, 16:319–376, 2005.

[4] F. Chyzak, A. Quadrat, and D. Robertz. OREMODULES: A symbolic package for the study of multidimensional linear systems. In *Applications of Time-Delay Systems*, volume 352 of *LNCIS*, pages 233–264. Springer, 2007.

[5] T. Cluzeau, V. Dolean, F. Nataf, A. Quadrat, Symbolic methods for developing new domain decomposition algorithms, INRIA Tehnical Report RR-7953, 2012. http://hal.inria.fr/hal-00694468

[6] A. D. Cox, J. Little, and D. O'Shea. *Using Algebraic Geometry*, volume 185 of *Graduate Texts in Mathematics*. Springer, second edition, 2005.

[7] V. Dolean and F. Nataf. A new domain decomposition method for the compressible Euler equations. *M2AN Math. Model. Numer. Anal.*, 40(4):689–703, 2006.

[8] V. Dolean, F. Nataf, and G. Rapin. Deriving a new domain decomposition method for the Stokes equations using the Smith factorization. *Math. Comp.*, 78(266):789–814, 2009.

[9] Ch. Farhat and F.-X. Roux. A Method of Finite Element Tearing and Interconnecting and its Parallel Solution Algorithm. *Internat. J. Numer. Methods Engrg.*, 32:1205–1227, 1991.

[10] P. Gosselet and C. Rey. Non-overlapping domain decomposition methods in structural mechanics. *Arch. Comput. Methods Engrg.*, 13(4):515–572, 2006.

[11] J. Mandel. Balancing domain decomposition. *Comm. on Applied Numerical Methods*, 9:233–241, 1992.

[12] F. Nataf. A new approach to perfectly matched layers for the linearized Euler system. *J. Comput. Phys.*, 214(2):757–772, 2006.

[13] L.F. Pavarino and O.B. Widlund. Balancing Neumann-Neumann methods for incompressible Stokes equations. *Comm. Pure Appl. Math.*, 55:302–335, 2002.

[14] A. Pechstein and J. Schöberl. Tangential-displacement and normal-normal-stress continuous mixed finite elements for elasticity. *Math. Models Methods Appl. Sci.*, 21(8):1761–1782, 2011.

[15] J.J. Rotman. *An Introduction to Homological Algebra*. Springer, second edition, 2009.

[16] P. Le Tallec, J. Mandel, and M. Vidrascu. A Neumann-Neumann Domain Decomposition Algorithm for Solving Plate and Shell Problems. *SIAM J. Numer. Anal.*, 35:836–867, 1998.

[17] J.T. Wloka, B. Rowley, and B. Lawruk. *Boundary Value Problems for Elliptic Systems*. Cambridge University Press, Cambridge, 1995.

Scalable Domain Decomposition Algorithms for Contact Problems: Theory, Numerical Experiments, and Real World Problems

Z. Dostál, T. Kozubek, T. Brzobohatý, A. Markopoulos, M. Sadowská, and V. Vondrák

Dept. of Appl. Math., VSB-Technical University of Ostrava, Czech Republic
zdenek.dostal@vsb.cz, tomas.kozubek@vsb.cz, tomas.brzobohaty@vsb.cz, alexandros.markopoulos@vsb.cz, marie.sadowska@vsb.cz, vit.vondrak@vsb.cz

Summary. We review our results related to the development of theoretically supported scalable algorithms for the solution of large scale contact problems of elasticity. The algorithms combine the Total FETI/BETI based domain decomposition method adapted to the solution of 2D and 3D multibody contact problems of elasticity, both frictionless and with friction, with our in a sense optimal algorithms for the solution of resulting quadratic programming and QPQC problems. Rather surprisingly, the theoretical results are qualitatively the same as the classical results on scalability of FETI/BETI for linear elliptic problems. The efficiency of the method is demonstrated by results of parallel numerical experiments for contact problems of linear elasticity discretized by more than 11 million variables in 3D and 40 million variables in 2D.

1 Introduction

Contact problems are in the heart of mechanical engineering. Solving large multibody contact problems of linear elastostatics is complicated by the inequality boundary conditions, which make them strongly non-linear, and, if the system of bodies includes "floating" bodies, by the positive semi-definite stiffness matrices resulting from the discretization of such bodies. Observing that the classical Dirichlet and Neumann boundary conditions are known only after the solution has been found, it is natural to assume the solution of contact problems to be more costly than the solution of a related linear problem with the classical boundary conditions. Since the cost of the solution of any problem increases at least linearly with the number of the unknowns, it follows that the development of a scalable algorithm for contact problems is a challenging task which requires to identify the contact interface in a sense for free.

The first promising results, at least for the frictionless problems, were obtained by the researchers who tried to modify the methods that were known to be scalable

R. Bank et al. (eds.), *Domain Decomposition Methods in Science and Engineering XX*, Lecture Notes in Computational Science and Engineering 91,
DOI 10.1007/978-3-642-35275-1_4,

for linear problems, in particular multigrid and domain decomposition. Experimental evidence of scalability was achieved with the monotonic multigrid (see [11] and the references therein). In spite of these nice results, the necessity to keep the coarse grid away from the contact interface prevented the authors to prove the optimality results similar to the classical results for linear problems. However, such result was obtained by Schöberl who has developed an approximate variant of the projection method using a domain decomposition preconditioner and a linear multigrid solver on the interior nodes. An experimental evidence of scalability for the frictionless problems was presented by Avery and Farhat [1]. The point of this paper is to report our optimality results for contact problems of linear elasticity, both frictionless and with friction.

The results are based on a combination of several ingredients. The first one is the application of the TFETI (Total FETI) [8] or TBETI (Total BETI) [14] methods, variants of the duality based domain decomposition methods introduced by Farhat and Roux [9] (finite elements) and Langer and Steinbach [13] (boundary elements). Since the TFETI/TBETI methods treat all the subdomains as "floating", the kernels of the stiffness matrices of the subdomains are a priori known. This makes the method very flexible and simplifies implementation of the multiplication of a vector by a generalized inverse of the stiffness matrix. As any duality based method, TFETI/TBETI reduces general inequality constraints to special separable ones.

The second ingredient is the "natural coarse grid preconditioning" introduced for linear problems by Farhat, Mandel, and Roux [10] and Langer and Steinbach [13]. This preconditioned cost function has the spectrum of the Hessian confined to a positive interval independent of the discretization parameter h and the decomposition parameter H provided the ratio H/h is uniformly bounded. Since our preconditioning uses a projector to the subspace with the solution, it follows that its application to the solution of variational inequalities does not turn the separable constraints into general constraints and can be interpreted as a variant of the multigrid method with the coarse grid on the interface. This unique feature, as compared with the standard multigrid preconditioning for the primal problem, reduces the development of scalable algorithms for the solution of variational inequalities to the solution of bound and equality constrained quadratic programming or QPQC (quadratic programming with quadratic constraints) problems with the rate of convergence in terms of bounds on the spectrum.

The resulting QP and QPQC problems, arising in the solution of the frictionless contact problems and the problems with the Tresca friction (an auxiliary problem for Coulomb friction), respectively, are solved by our algorithms with the rate of convergence in terms of the bounds on the spectrum, the third ingredient of our development (see [7]). Putting the three ingredients together with a few simple observations, we get theoretically supported algorithms for contact problems. The theoretical results are illustrated by the results of numerical experiments which show that both numerical and parallel scalability can be observed in practice. Finally we report the solutions of some real world problems. More details can be found in Dostál et al. [3–5], and Sadowská et al. [14].

2 Dual Formulation of Frictionless Contact Problems

To simplify our presentation, let us assume that the bodies are assembled from N_s subdomains $\Omega^{(s)}$ which are "glued" together by suitable equality constraints. After the standard finite element discretization, the equilibrium of the system is described as a solution u of the problem

$$\min J(v) \quad \text{subject to} \quad \sum_{s=1}^{N_s} B_N^{(s)} v^{(s)} \le g_N \quad \text{and} \quad \sum_{s=1}^{N_s} B_E^{(s)} v^{(s)} = o, \tag{1}$$

where o denotes the zero vector and $J(v)$ is the energy functional defined by

$$J(v) = \sum_{s=1}^{N_s} \frac{1}{2} {v^{(s)}}^T K^{(s)} v^{(s)} - {v^{(s)}}^T f^{(s)},$$

$v^{(s)}$ and $f^{(s)}$ denote the admissible subdomain displacements and the subdomain vector of prescribed forces, $K^{(s)}$ is the subdomain stiffness matrix, $B_N^{(s)} \in \mathbb{R}^{m_C \times n}$ and $B_E^{(s)} \in \mathbb{R}^{m_E \times n}$ are the blocks of the matrix $B = \left[B_N^T, B_E^T\right]^T$ that correspond to $\Omega^{(s)}$, and g_N is a vector collecting the normal gaps between the bodies in the reference configuration. The matrix B_N and the vector g_N arise from the nodal or mortar description of the non-penetration conditions, while B_E describes the "gluing" of the subdomains into the bodies and the Dirichlet boundary conditions. Recall that if the problem is discretized by the TBETI method, then we get the potential energy minimization problem of the very same structure as (1), where all the objects correspond only to the boundaries $\Gamma^{(s)}$ of $\Omega^{(s)}$ except the term with the prescribed volume forces (if there is some); see [14] for more details. By contrast with TFETI, when the matrices $K^{(s)}$ are sparse, in the case of TBETI these are fully populated.

To simplify the presentation of basic ideas, we can describe the equilibrium in terms of the global stiffness matrix K, the vector of global displacements u, and the vector of global loads f. In the TFETI/TBETI methods, we have

$$K = \operatorname{diag}(K^{(1)}, \ldots, K^{(N_s)}), \quad u = \begin{bmatrix} u^{(1)} \\ \vdots \\ u^{(N_s)} \end{bmatrix}, \quad \text{and} \quad \mathrm{f} = \begin{bmatrix} f^{(1)} \\ \vdots \\ f^{(N_s)} \end{bmatrix},$$

where $K^{(s)}$, $s = 1, \ldots, N_s$, is a positive semidefinite matrix. The energy function reads

$$j(v) = \frac{1}{2} v^T K v - f^T v$$

and the vector of global displacements u solves

$$\min j(v) \quad \text{s.t.} \quad B_N v \le g_N \quad \text{and} \quad B_E v = o.$$

Alternatively, the global equilibrium may be described by the Karush–Kuhn–Tucker conditions (see, e.g., [6])

$$Ku = f - B^T\lambda, \quad \lambda_N \geq o, \quad \lambda^T(Bu - g) = o, \tag{2}$$

where $g = \left[g_N^T, o^T\right]^T$ and $\lambda = \left[\lambda_N^T, \lambda_E^T\right]^T$ denotes the vector of Lagrange multipliers which may be interpreted as the reaction forces. The problem (2) differs from the linear problem by the non-negativity constraint on the components of reaction forces λ_N and by the complementarity condition.

We can use the first equation of (2) to eliminate the displacements. We shall get the problem to find

$$\min \Theta(\lambda) \quad \text{s.t.} \quad \lambda_N \geq o \quad \text{and} \quad R^T(f - B^T\lambda) = o, \tag{3}$$

where

$$\Theta(\lambda) = \frac{1}{2}\lambda^T BK^+B^T\lambda - \lambda^T(BK^+f - g) + \frac{1}{2}fK^+f, \tag{4}$$

K^+ denotes a generalized inverse that satisfies $KK^+K = K$, and R denotes the full rank matrix whose columns span the kernel of K. The action of K^+ can be evaluated at the cost comparable with that of Cholesky's decomposition applied to the regularized K (see [2]). Denoting $\mathscr{F} = \|BK^+B^T\|$,

$$F = \mathscr{F}^{-1}BK^+B^T, \quad e = SR^Tf, \quad G = SR^TB^T, \quad \widetilde{d} = \mathscr{F}^{-1}(BK^\dagger f - g),$$

with S denoting a nonsingular matrix that defines the orthonormalization of the rows of R^TB^T, we can modify (3) to

$$\min \widetilde{\theta}(\lambda) \quad \text{s.t.} \quad \lambda_N \geq 0 \quad \text{and} \quad G\lambda = e, \tag{5}$$

where

$$\widetilde{\theta}(\lambda) = \frac{1}{2}\lambda^T F\lambda - \lambda^T\,\widetilde{d}. \tag{6}$$

Our next step is to replace the equality constraint in (5) by a homogeneous one. To this end, it is enough to find any $\widetilde{\lambda}$ such that

$$G\widetilde{\lambda} = e,$$

denote $\lambda = \mu + \widetilde{\lambda}$, and substitute into (5). We get

$$\widetilde{\theta}(\lambda) = \frac{1}{2}\mu^T F\mu - \mu^T(\widetilde{d} - F\widetilde{\lambda}) + const.$$

After returning to the old notation, problem (5) is reduced to

$$\min \frac{1}{2}\lambda^T F\lambda - \lambda^T\, d \quad \text{s.t.} \quad G\lambda = o \quad \text{and} \quad \lambda_N \geq \ell_N \tag{7}$$

with $\ell = -\widetilde{\lambda}$ and $d = \widetilde{d} - F\widetilde{\lambda}$. Since G has orthonormal rows, we can use the least square solution

$$\widetilde{\lambda} = G^Te. \tag{8}$$

3 Dual Formulation of Contact Problems with Tresca Friction

If the Tresca friction is prescribed on the contact interface, then the equilibrium of the system is described as a solution u of the problem

$$\min J_T(v) \quad \text{subject to} \quad \sum_{s=1}^{N_s} B_N^{(s)} v^{(s)} \le g_N \quad \text{and} \quad \sum_{s=1}^{N_s} B_E^{(s)} v^{(s)} = o, \tag{9}$$

where $J_T(v)$ is the energy functional defined by

$$J_T(v) = J(v) + j(v), \quad j(v) = \sum_{i=1}^{m_C} \Psi_i \|T_i u\|,$$

Ψ_i denotes an a priori defined slip bound at node i, and $T_i u$ denotes the jump of the tangential displacement due to the displacement u. Using the standard procedure to modify the non-differentiable term j (see [3, 5]), we get

$$j(v) = \sum_{i=1}^{m_C} \Psi_i \|T_i u\| = \sum_{i=1}^{m_C} \max_{\|\tau_i\| \le \Psi_i} \tau_i^T T_i u,$$

where τ_i can be considered as Lagrange multipliers. We assume that B_N, B_E, and T are full rank matrices.

Let $\overline{d}$ denote the spatial dimension and let us introduce the Lagrangian with three types of Lagrange multipliers, namely $\lambda_N \in \mathbb{R}^{m_C}$ associated with the non-interpenetration condition, $\lambda_E \in \mathbb{R}^{m_E}$ associated with the "gluing" and prescribed displacements, and

$$\tau = [\tau_1^T, \tau_2^T, \ldots, \tau_{m_C}^T]^T \in \mathbb{R}^{(\overline{d}-1)m_C}$$

which regularizes the non-differentiability. The Lagrangian associated with problem (1) reads

$$L(u, \lambda_N, \lambda_E, \tau) = J(u) + \tau^T T u + \lambda_N^T (B_N u - c_N) + \lambda_E^T (B_E u - c_E). \tag{10}$$

Using the convexity of the cost function and constraints, we can use the classical duality theory [6] to reformulate problem (9) to get

$$\min_u \sup_{\substack{\lambda_E \in \mathbb{R}^{m_E},\ \lambda_N \ge \mathbf{o} \\ \|\tau_i\| \le \Psi_i,\ i=1,\ldots,m_C}} L(u, \lambda_N, \lambda_E, \tau) = \max_{\substack{\lambda_E \in \mathbb{R}^{m_E},\ \lambda_N \ge o \\ \|\tau_i\| \le \Psi_i,\ i=1,\ldots,m_C}} \min_u L(u, \lambda_N, \lambda_E, \tau).$$

To simplify the notation, we denote

$$\lambda = \begin{bmatrix} \lambda_E \\ \lambda_N \\ \tau \end{bmatrix}, \quad B = \begin{bmatrix} B_E \\ B_N \\ T \end{bmatrix}, \quad c = \begin{bmatrix} c_E \\ c_N \\ o \end{bmatrix},$$

and

$$\Lambda(\Psi) = \left\{(\lambda_E^T, \lambda_N^T, \tau^T)^T \in \mathbb{R}^{m_E+\bar{d}m_C} : \ \lambda_N \geq o, \|\tau_i\| \leq \Psi_i, \ i = 1, \ldots, m_C\right\},$$

so that we can write the Lagrangian briefly as

$$L(u, \lambda) = \frac{1}{2} u^T K u - f^T u + \lambda^T (Bu - c)$$

and problem (9) is equivalent to the saddle point problem

$$L(\widehat{u}, \widehat{\lambda}) = \max_{\lambda \in \Lambda(\Psi)} \min_{u} L(u, \lambda). \tag{11}$$

Similarly to the frictionless case, we eliminate the primal variables from (11) and carry out the homogenization to reduce the minimization problem to

$$\min \frac{1}{2} \lambda^T F \lambda - \lambda^T d \quad \text{s.t.} \quad G\lambda = o \quad \text{and} \quad \lambda \in \Lambda(\Psi) \tag{12}$$

with the notation of Sect. 2. Notice that we minimize exactly the same type of the cost function as in the frictionless case, but with some additional quadratic constraints.

4 Preconditioning by Projector

Our final step is based on the observation that both the frictionless contact problem and the contact problem with Tresca friction are equivalent to

$$\min \theta(\lambda) \quad \text{s.t.} \quad \lambda \in \Omega, \tag{13}$$

where

$$\theta(\lambda) = \frac{1}{2} \lambda^T (PFP + \overline{\rho} Q) \lambda - \lambda^T P d, \quad Q = G^T (G G^T)^{-1} G, \quad P = I - Q,$$

$\overline{\rho} > 0$, and $\Omega = \{\lambda : G\lambda = o \text{ and } \lambda_N \geq o\}$ (without friction) or $\Omega = \{\lambda : G\lambda = o$ and $\lambda \in \Lambda(\Psi)\}$ (Tresca). A good choice of the regularization parameter is given by

$$\overline{\rho} = \|PFP\|,$$

as this is the largest value for which

$$\|PFP\| \geq \|PFP + \overline{\rho} Q\|.$$

Problem (13) turns out to be a suitable starting point for development of an efficient algorithm for variational inequalities due to the following classical estimates [10] of the extreme eigenvalues.

Theorem 1. *If the decompositions and the discretizations of given contact problems are sufficiently regular, then there are constants $C_1 > 0$ and $C_2 > 0$ independent of the discretization parameter h and the decomposition parameter H such that*

$$C_1 \frac{h}{H} \ \leq \ \lambda_{\min}(PFP|\mathrm{Im}P) \quad \textit{and} \quad \lambda_{\max}(PFP|\mathrm{Im}P) = \|PFP\| \ \leq \ C_2, \tag{14}$$

where $\lambda_{\min}$ and $\lambda_{\max}$ denote the extremal eigenvalues of the corresponding matrices.

5 Optimality

Theorem 1 states that if we fix the regularization parameter $\overline{\rho}$ and keep H/h uniformly bounded, then problem (13) resulting from the application of various discretizations and decompositions has the spectrum of the Hessian matrices confined to a positive interval. It follows that to develop a scalable algorithm for the contact problems, it is enough to find an algorithm that is able to find an approximate solution of (13) in a number of matrix–vector multiplications uniformly bounded in terms of bounds on the spectrum of the cost function.

Here we propose to use SMALSE (semi-monotonic augmented Lagrangian method for separable and equality constraints), our variant of the augmented Lagrangian method [7]. SMALSE enforces the equality constraints by the Lagrange multipliers generated in the outer loop, while the auxiliary QPQC problems with separable constraints are solved approximately in the inner loop by the MPGP algorithm proposed by Dostál and Kozubek [7]. MPGP is an active set based algorithm which uses the conjugate gradient method to explore the current face, the fixed steplength gradient projection to change the active set, and the adaptive precision control for the solution of auxiliary linear problems. The unique feature of SMALSE with the inner loop implemented by MPGP when used to (13) is the bound on the number of iterations whose cost is proportional to the number of variables, so that it can return an approximate solution for the cost proportional to the number of variables. It follows that SMALSE/MPGP is a scalable algorithm for the solution of (13) provided the cost of decomposition of K and application of the projectors P and Q is not too large.

Theorem 2. *If the decompositions and the discretizations of a given contact problem are sufficiently regular, then there is a constant $C > 0$ independent of the discretization parameter h and the decomposition parameter H such that the algorithm SMALSE/MPGP (or SMALBE/MPRGP for the frictionless problems) with fixed parameters specified in [7] can find the solution of (13) in a number of iterations bounded by C provided the initial approximation satisfies*

$$\|\lambda^0\| \le c\|Pd\|,$$

where $c > 0$ is an a priori chosen constant.

6 Numerical Experiments

The algorithms reported in this paper were implemented into our MatSol software [12] and tested with the aim to verify their optimality and capability to solve the real world problems.

6.1 Scalability of TFETI: 2D Cantilever Beams with Tresca Friction

We first tested the scalability on a 2D problem of Fig. 1 with varying discretizations and decompositions using structured grids. We kept the ratio H/h of the

decomposition and the discretization parameters approximately constant so that the assumptions of Theorem 1 were satisfied.

The results of computations carried out to the relative precision 10^{-4} are in Table 1. We can observe that the number of matrix–vector multiplications varies only mildly with the increasing dimension of the problem in agreement with the theory. We conclude that the scalability can be observed in practice.

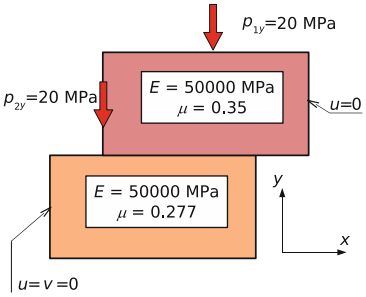


Fig. 1. Geometry of 2D cantilever beams

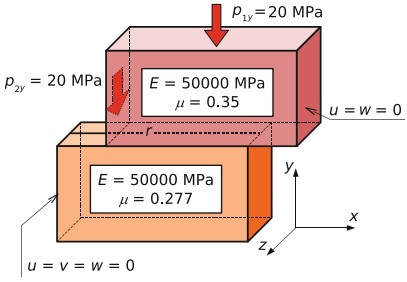


Fig. 2. Geometry of 3D cantilever beams

Table 1. Numerical scalability of TFETI: 2D cantilever beams.

Number of subdomains	1936	4096	7744
Number of CPUs	48	48	48
Primal variables	10,071,072	21,307,392	40,284,288
Dual variables	384,473	817,793	1,551,089
Null space dimension	5808	12,288	23,232
SMALSE-M iterations	8	8	8
Hessian multiplications	119	134	180
Solution time [s]	839	1665	7825

6.2 Scalability of TFETI/TBETI: 3D Cantilever Beams with Tresca Friction

The second problem was a 3D alternative to the previous example (see Fig. 2). The results of computations carried out for both TFETI and TBETI methods are in Tables 2 and 3, respectively. We can see that the number of matrix–vector multiplications again varies only mildly with the increasing problem size as predicted by the theory.

Table 2. Numerical scalability of TFETI: 3D cantilever beams.

Number of subdomains	108	500	1372	2916
Number of CPUs	48	48	48	48
Primal variables	431,244	1,996,500	5,478,396	11,643,588
Dual variables	88,601	444,927	1,261,493	2,728,955
Null space dimension	648	3000	8232	17,496
SMALSE-M iterations	3	4	4	4
Hessian multiplications	78	97	93	119
Solution time [s]	60	374	1663	7745

Table 3. Numerical scalability of TBETI: 3D cantilever beams.

Number of subdomains	108	500	1372	2916
Number of CPUs	48	48	48	48
Primal variables	195,045	903,000	2,477,830	5,266,300
Dual variables	88,601	444,927	1,261,493	2,728,955
Null space dimension	648	3000	8232	17,496
SMALSE-M iterations	7	8	9	9
Hessian multiplications	160	161	160	260
Solution time [s]	46	301	2211	7949

6.3 Applications of TFETI/TBETI to Real World Problems

We have also tested our algorithms on real world problems. First we consider the analysis of the stress in the roller bearings of Fig. 3. The problem is difficult because it consists of 73 bodies in mutual contact and only one is fixed in space. The solution of the problem discretized by 2,730,000/459,800 primal/dual variables and decomposed into 700 subdomains required 4,270 matrix–vector multiplications. The von Mises stress distribution is in Fig. 3.

Second we consider the analysis of the yielding clamp connection of steel arched supports depicted in Fig. 4. This type of construction is used to support the mining openings. It is a typical multibody contact, where the yielding connection plays the role of the mechanical protection against destruction, i.e., against the total deformation of the supporting arches. We consider contact with the Coulomb friction, where the coefficient of friction was $\mathscr{F} = 0.5$. The problem was decomposed into 250 subdomains using METIS and discretized by 1,592,853 and 216,604 primal and dual variables, respectively. The total displacements for both TFETI and TBETI are depicted in Fig. 4. The solution required 1,922 matrix-vector multiplications.

7 Comments and Conclusions

The TFETI method turns out to be a powerful engine for the solution of contact problems of elasticity. The results of numerical experiments comply with the theoretical results and indicate high efficiency of the method reported here. Future research will include adaptation of the standard preconditioning strategies.

Fig. 3. Frictionless roller bearing of wind generator

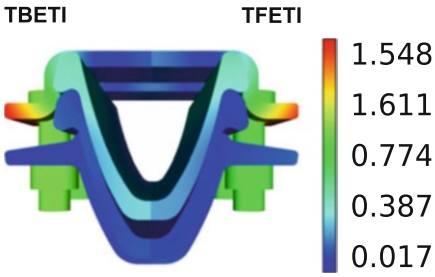


Fig. 4. Steel support with Coulomb friction

Acknowledgments This research has been supported by the grants GA CR No. 201/07/0294 and ME CR No. MSM6198910027.

Bibliography

[1] Philip Avery and Charbel Farhat. The FETI family of domain decomposition methods for inequality-constrained quadratic programming: application to contact problems with conforming and nonconforming interfaces. *Comput. Methods Appl. Mech. Engrg.*, 198(21–26):1673–1683, 2009. ISSN 0045-7825. doi: 10.1016/j.cma.2008.12.014. URL http://dx.doi.org/10.1016/j.cma.2008.12.014.

[2] T Brzobohatý, Z Dostál, T Kozubek, P Kovář, and A Markopoulos. Cholesky decomposition with fixing nodes to stable computation of a generalized inverse of the stiffness matrix of a floating structure. Accepted, 2011.

[3] Z. Dostál, T. Kozubek, P. Horyl, T. Brzobohatý, and A. Markopoulos. A scalable TFETI algorithm for two-dimensional multibody contact problems with friction. *J. Comput. Appl. Math.*, 235(2):403–418, 2010. ISSN 0377-0427. doi: 10.1016/j.cam.2010.05.042. URL http://dx.doi.org/10.1016/j.cam.2010.05.042.

[4] Z. Dostál, T. Kozubek, V. Vondrák, T. Brzobohatý, and A. Markopoulos. Scalable TFETI algorithm for the solution of multibody contact problems of elasticity. *Internat. J. Numer. Methods Engrg.*, 82(11):1384–1405, 2010. ISSN 0029-5981.

[5] Z. Dostál, T. Kozubek, A. Markopoulos, T. Brzobohatý, V. Vondrák, and P. Horyl. A theoretically supported scalable tfeti algorithm for the solution of multibody 3d contact problems with friction. In Press, 2011.

[6] Zdeněk Dostál. *Optimal quadratic programming algorithms*, volume 23 of *Springer Optimization and Its Applications*. Springer, New York, 2009. ISBN 978-0-387-84805-1. With applications to variational inequalities.

[7] Zdeněk Dostál and Tomáš Kozubek. An optimal algorithm and superrelaxation for minimization of a quadratic function subject to separable convex constraints with applications. *Mathematical Programming*, pages 1–26, 2011. ISSN 0025-5610. doi: 10.1007/s10107-011-0454-2.

[8] Zdeněk Dostál, David Horák, and Radek Kučera. Total FETI—an easier implementable variant of the FETI method for numerical solution of elliptic PDE. *Comm. Numer. Methods Engrg.*, 22(12):1155–1162, 2006. ISSN 1069-8299. doi: 10.1002/cnm.881. URL http://dx.doi.org/10.1002/cnm.881.

[9] Charbel Farhat and François-Xavier Roux. A method of finite element tearing and interconnecting and its parallel solution algorithm. *Internat. J. Numer. Methods Engrg.*, 32(6):1205–1227, 1991. ISSN 1097-0207. doi: 10.1002/nme.1620320604. URL http://dx.doi.org/10.1002/nme.1620320604.

[10] Charbel Farhat, Jan Mandel, and François-Xavier Roux. Optimal convergence properties of the FETI domain decomposition method. *Comput. Methods Appl. Mech. Engrg.*, 115(3–4):365–385, 1994. ISSN 0045-7825. doi: 10.1016/0045-7825(94)90068-X. URL http://dx.doi.org/10.1016/0045-7825(94)90068-X.

[11] Ralf Kornhuber. *Adaptive monotone multigrid methods for nonlinear variational problems*. Advances in Numerical Mathematics. B. G. Teubner, Stuttgart, 1997. ISBN 3-519-02722-4.

[12] T. Kozubek, A. Markopoulos, T. Brzobohatý, R. Kučera, V. Vondrák, and Z. Dostál. Matsol - matlab efficient solvers for problems in engineering. "http://matsol.vsb.cz/", 2009.

[13] U. Langer and O. Steinbach. Boundary element tearing and interconnecting methods. *Computing*, 71(3):205–228, 2003. ISSN 0010-485X. doi: 10.1007/s00607-003-0018-2. URL http://dx.doi.org/10.1007/s00607-003-0018-2.

[14] M. Sadowská, Z. Dostál, T. Kozubek, A. Markopoulos, and J. Bouchala. Scalable total beti based solver for 3d multibody frictionless contact problems in mechanical engineering. *Eng. Anal. Bound. Elem.*, 35(3):330–341, 2011. ISSN 0955-7997. doi: 10.1016/j.enganabound.2010.09.015.

Robust Coarsening in Multiscale PDEs

Robert Scheichl[1]

Dept. Mathematical Sciences, University of Bath, Bath, UK. R.Scheichl@bath.ac.uk

1 Introduction

Consider a variationally–posed second–order elliptic boundary value problem

$$a(u,v) \equiv \int_\Omega \mathscr{A}(\mathbf{x})\,\nabla u\cdot\nabla v = \int_\Omega f(\mathbf{x})v(\mathbf{x}), \quad \text{for all } v\in H^1_0(\Omega), \tag{1}$$

with solution $u\in H^1_0(\Omega)$ and domain $\Omega\subset\mathbb{R}^d$, $d=2,3$, where the coefficient tensor $\mathscr{A}(\mathbf{x})$ is highly *heterogeneous* (possibly in a spatially complicated way). We assume that $\mathscr{A}(\mathbf{x})$ is symmetric, uniformly positive definite and mildly anisotropic, i.e. $\lambda_{\min}(\mathscr{A}(\mathbf{x}))\gtrsim\lambda_{\max}(\mathscr{A}(\mathbf{x}))$ uniformly in $\mathbf{x}$. We are particularly interested in the case when the *contrast* $\max_{\mathbf{x},\mathbf{y}\in\Omega}\lambda_{\max}(\mathscr{A}(\mathbf{x}))/\lambda_{\max}(\mathscr{A}(\mathbf{y}))$ is large. Many examples of this type arise in subsurface flow modelling or in material science. The space $H^1_0(\Omega)$ is the usual Sobolev space of functions with vanishing trace on $\partial\Omega$ and $f\in H^{-1}(\Omega)$. For simplicity we assume for the remainder that $\mathscr{A}(\mathbf{x})=\alpha(\mathbf{x})I$, i.e. a scalar diffusion coefficient.

Let $\mathscr{T}_h$ be a simplicial triangulation of Ω and let (1) be discretised in $V_h\subset H^1_0(\Omega)$, the space of continuous, piecewise linear FE functions with respect to $\mathscr{T}_h$ that vanish on $\partial\Omega$. For simplicity let $\mathscr{T}_h$ be quasi-uniform. The a-orthogonal projection of u to V_h is denoted by u_h. In the usual nodal basis $\{\varphi_i\}_{i=1}^n$ for V_h, the problem of finding u_h reduces to the $n\times n$ linear system

$$A\mathbf{u}=\mathbf{b} \tag{2}$$

with stiffness matrix $A=(a(\varphi_i,\varphi_j))_{i,j=1}^n$. Since the matrix A depends on α only through element averages, we can assume (w.l.o.g.) that α is piecewise constant with respect to $\mathscr{T}_h$. For simplicity we assume that α is piecewise constant with respect to some non-overlapping partitioning of Ω into open, connected Lipschitz polyhedra (polygons) $\{\mathscr{Y}_m\}_{m=1}^M$ and set $\alpha_m=\alpha|_{\mathscr{Y}_m}$.

Especially for $d=3$ and for problems where α varies on a small length scale $\varepsilon\ll\mathrm{diam}(\Omega)$, and thus the mesh size h needs to be very fine, multilevel iterative solvers (multigrid, domain decomposition, etc.) are usually essential to solve

R. Bank et al. (eds.), *Domain Decomposition Methods in Science and Engineering XX*, Lecture Notes in Computational Science and Engineering 91,
DOI 10.1007/978-3-642-35275-1_5,

this problem efficiently. Their scalability and robustness with respect to mesh refinement, as well as other discretisation parameters has been studied extensively. Here we will focus on their robustness with respect to coefficient variation. We will show that coefficient robustness is inherently linked to a judicious choice of coarse space V_H (related to some coarse mesh $\mathscr{T}_H$ with resolution H). If $\varepsilon \gtrsim H$ and if we can choose a coarse mesh such that all coefficient jumps are aligned with the mesh, then the coefficient robustness of standard coarse spaces has been analysed in the 1990s (cf. [3, 4, 10, 16, 21, 22, 25] and the references therein). For certain methods the robustness may depend on the quasi-monotonicity of the coefficient with respect to the coarse mesh (in the sense of [3]). Substructuring-type ("exotic") coarse spaces are usually used to achieve uniform coefficient robustness. A certain amount of robustness can be recovered for standard piecewise linear coarse spaces by using the multilevel solver as a preconditioner within CG (e.g. [24]). The key tool in all these analyses is the weighted L_2–projection of Bramble and Xu [1]. It requires a piecewise constant weight with respect to the coarse mesh, an assumption that is often far too stringent in real applications. We want to move away from this and crucially here make no assumptions that the underlying coarse grids resolve the coefficients.

A lot of effort in the last 25 years has gone into the development of algebraic methods to construct coarse spaces, such as algebraic multigrid (AMG), rather than analytic/geometric ones. It has been confirmed numerically that AMG methods are in practice robust to coefficient variation when applied to (2) (i.e. the number of iterations is unaffected), and they are therefore extremely popular. However, they are built on several heuristics and so a rigorous analysis of their coefficient-robustness is difficult (see [22] for a review of existing theoretical results). Nevertheless, the key principle of these algebraic coarse spaces, namely energy minimisation [11], also underlies many other coarse spaces. To obtain rigorous coefficient–independent convergence results we will need to work in the following energy and weighted L_2-norms on $D \subset \Omega$,

$$\|v\|_{a,D} = \int_D \alpha |\nabla v|^2 \qquad \text{and} \qquad \|v\|_{0,\alpha,D} = \int_D \alpha v^2 ,$$

respectively. When $D = \Omega$ we will usually not specify the domain explicitly.

A convenient framework to analyse most multilevel methods is the Schwarz or subspace correction framework [21, 23]. We restrict attention to the two-level overlapping additive Schwarz method and focus on the robustness of various coarse spaces for this method. We review some recent papers on the topic mainly by the author (jointly with co-workers), as well as by Efendiev et al. All the results apply immediately also to multiplicative, hybrid and non-overlapping versions of the Schwarz method (see [9, 19] for some explicit comments). Many of the results can be extended to a multilevel theory [5, 19].

2 Schwarz Framework and Abstract Coarse Spaces

Let us assume that $\{\Omega_k\}_{k=1}^K$ is an overlapping partitioning of Ω and let Ω_k° be the overlap of subdomain Ω_k, i.e. the set of points $\mathbf{x} \in \Omega_k$ that are contained in at least one

other subdomain. We assume that $\mathscr{T}_h$ is aligned with this partitioning. Furthermore, let $\{\chi_k\}_{k=1}^K \subset V_h$ be an arbitrary partition of unity (POU) of FE functions subordinate to $\{\Omega_k\}_{k=1}^K$ such that $\|\chi_k\|_\infty \lesssim 1$ and $\|\nabla\chi_k\|_\infty \leq \delta_k^{-1}$, for all $k = 1,\ldots,K$. Note that (due to quasi-uniformity of $\mathscr{T}_h$) we always have $\delta_k \gtrsim h$, and there is a partition of unity such that δ_k is proportional to the (minimal) width of Ω_k°. We assume as usual that each point $\mathbf{x} \in \Omega$ is contained in at most N_0 subdomains (*finite covering*).

We associate with each Ω_k the space $V_k = \{v \in V_h : \mathrm{Supp}(v) \subset \overline{\Omega}_k\}$ and assume that we have an additional *coarse space*

$$V_0 = V_H = \mathrm{span}\{\Phi_j \in V_h : j = 1,\ldots,N\} \subset V_h\,.$$

Let $\omega_j = \mathrm{interior}(\mathrm{Supp}(\Phi_j))$ and set $H_j = \mathrm{diam}(\omega_j)$. Then $H = \max_j H_j$ is the coarse mesh size associated with V_H.

The two-level additive Schwarz preconditioner is now simply

$$M_{\mathrm{AS}}^{-1} = R_0^T A_0^{-1} R_0 \;+\; \sum_{k=1}^K R_k^T A_k^{-1} R_k \quad \text{with} \quad A_k = R_k A R_k^T\,.$$

R_k is the matrix representation of a restriction operator from V to V_k: the simple injection operator for $k \geq 1$, and for $k = 0$ induced by the coarse space basis $\{\Phi_j\}_{j=1}^N$ so that the coarse space stiffness matrix is $A_0 = (a(\Phi_j, \Phi_\ell))_{j,\ell}^N$.

The following result can be proved in the same way as [19, Theorem 2.5]. Since it is instructive, we give an outline of the proof.

Theorem 1. *If there exists an operator $\Pi : V_h \to V_0$ such that for all $v \in V_h$*

$$\|\Pi v\|_a^2 \;\leq\; C_1\,\|v\|_a^2 \quad \textit{and} \quad \sum_{k=1}^K \|(v - \Pi v)\nabla\chi_k\|_{0,\alpha}^2 \;\leq\; C_2\,\|v\|_a^2, \tag{3}$$

then $\kappa(M_{\mathrm{AS}}^{-1}A) \lesssim C_1 + C_2$. The hidden constant depends on N_0.

Proof. Let $v_0 = \Pi v$ be such that (3) holds and choose $v_k = I_h(\chi_k(v - v_0))$, where I_h is the standard nodal interpolant on V_h. This interpolant is stable for all piecewise quadratic functions in the energy norm and in the weighted L_2-norm (independently of α) (cf. [19, Lemma 2.3]), and so we get

$$\begin{aligned}\sum_{k=0}^K \|v_k\|_a^2 \;&\lesssim\; \|v_0\|_a^2 + \sum_{k=1}^K \|\chi_k(v - v_0)\|_a^2 \\ &\lesssim\; \|v_0\|_a^2 + \sum_{k=1}^K \|\chi_k\|_\infty^2 \|v - v_0\|_{a,\Omega_k}^2 + \|(v - v_0)\nabla\chi_k\|_{0,\alpha}^2\,.\end{aligned}$$

Now, the boundedness of the POU functions, the finite cover assumption, as well as (3) lead to the stability estimate $\sum_{k=0}^K \|v_k\|_a^2 \lesssim (C_1 + C_2)\|v\|_a^2$. Since $v = \sum_{k=0}^K v_k$, the result follows from the abstract Schwarz theory (cf. [21]).

This result shows the importance of the choice of coarse space. Provided we have a good coarse space approximation in the weighted L_2-norm that is moreover stable in the energy norm, independently of variations in α, then the bound on the condition number for two-level additive Schwarz is also robust with respect to these variations. Note that it is crucial to use the weighted L_2 and the energy norm here to achieve

coefficient-robustness, and that we only require weak L_2–approximation in regions where $\nabla\chi_k \neq 0$.

Several approaches have been studied in [2, 5–9, 17–19] to provide constants in (3) that are independent of α (or at least of the contrast in α) for various coarse spaces. However, in most cases the constants are not independent of $\frac{H}{\varepsilon}$, where ε is the minimal length scale at which α varies in the regions where $\nabla\chi_k \neq 0$. So unfortunately in general, to be also independent of $\frac{H}{\varepsilon}$, restrictions on the coarse mesh size are needed, at least locally.

Let us discuss the assumptions (3) a bit further. Let $\Pi v = \sum_j f_j(v)\Phi_j$, where $f_j : V_h \to \mathbb{R}$ is a suitable functional. Then

$$\|\Pi v\|_a \;=\; \big\|\textstyle\sum_j f_j(v)\Phi_j\big\|_a \;\leq\; \textstyle\sum_j |f_j(v)|\|\Phi_j\|_a\,.$$

We see that a set of coarse basis functions with bounded energy (independent of α) is beneficial. The first approaches in [8, 9, 17] attacked this target directly and aimed at bounding $\|\Phi_j\|_a$. In that case, it suffices to use the standard quasi-interpolant. Alternatively, a weighted quasi-interpolant with $f_j(v) = \int_{\omega_j}\alpha v / \int_{\omega_j}\alpha$ can be used. For certain (locally quasi-monotone) coefficients α this leads to a constant C_1 that is independent of the contrast in α, even if the energy of the basis functions is not bounded (see below).

Similar comments can be made about the second assumption in (3). Note that

$$\|(v-\Pi v)\nabla\chi_k\|^2_{0,\alpha} \;\leq\; \begin{cases} \|\alpha|\nabla\chi_k|^2\|_\infty \|v-\Pi v\|^2_{0,\Omega_k^\circ}\,, & \text{or} \\ \|\nabla\chi_k\|^2_\infty \|v-\Pi v\|^2_{0,\alpha,\Omega_k^\circ}\,. \end{cases}$$

We can either try to choose a partition of unity $\{\chi_k\}$ such that $\|\alpha|\nabla\chi_k|^2\|_\infty$ is bounded independently of α, which is again related to energy minimisation, or we can try to bound $\|v-\Pi v\|_{0,\alpha,\Omega_k^\circ}$ directly. As above, it is possible for certain (locally quasi-monotone) coefficients to achieve this and to obtain a constant C_2 that does not depend on the contrast in α (see below).

When the coefficient is not locally quasi-monotone, then it is in general necessary to enrich the coarse space, by either refining the coarse mesh locally, or by choosing more than one basis function per subdomain Ω_k, with the key tool to achieve coarse space robustness being again energy minimisation.

To highlight some of the key issues we will use a number of representative model problems shown in Fig. 1. For the rest of the paper, we will only focus on cases, such as Fig. 1c–h, where it is impossible or impractical that the subdomains $\{\Omega_k\}$ and the supports $\{\omega_j\}$ of the coarse basis functions resolve the coefficient jumps. The resolved cases in Fig. 1a, b have already been studied extensively, see e.g. [3, 4, 10, 16, 21, 22, 24, 25].

3 Analysis of Coefficient–Robustness

We present three possible approaches to try and prove coefficient robustness rigorously and thus to design robust coarse spaces. For simplicity, we assume that for each $j = 1,\ldots,N$, there exists a $k = 1,\ldots,K$ such that $\omega_j \subset \Omega_k$.

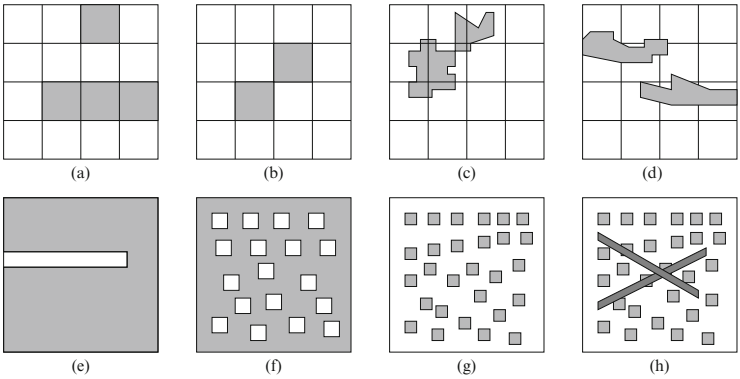


Fig. 1. Typical coefficient distributions (**a**) resolved; (**b**) not quasi-monotone; (**c**) neither quasi-monotone nor resolved; (**d**) channelised; (**e**) flow barriers; (**f**) low permeability inclusions; (**g**) high permeability inclusions; (**h**) high permeability inclusions and channels

3.1 Standard Quasi-interpolant and Energy Minimisation

The first approach makes use of the standard quasi-interpolant

$$\Pi v \;=\; \sum_{j=1}^{N} \overline{v}_{\omega_j} \Phi_j, \qquad \text{where} \qquad \overline{v}_{\omega_j} \;=\; \tfrac{1}{|\omega_j|} \textstyle\int_{\omega_j} v\,.$$

Let $\{\Phi_j\}_{j=1}^N$ be a set of bounded coarse basis functions that form a partition of unity, except in a boundary layer of width $\mathscr{O}(H)$ near $\partial\Omega$. Since each support $\omega_j \subset \Omega_k$, for some k, the supports have finite overlap. The constants C_1 and C_2 can now be bounded independent of the contrast in α, if either

$$\gamma_2(\alpha,\{\Phi_j\}) = \max_{j=1}^{N} H_j^{2-d}\,\|\Phi_j\|_a^2 \quad \text{and} \quad \gamma_\infty(\alpha,\{\chi_k\}) = \max_{k=1}^{K} \delta_k^2\,\|\alpha^{1/2}\nabla\chi_k\|_\infty^2$$

(the so-called *coarse space and partitioning robustness indicators*) can be bounded independent of α, for some choice of the partition of unity $\{\chi_k\}_{k=1}^K$ subordinate to $\{\Omega_k\}_{k=1}^K$ (cf. [8]), or if $\gamma_\infty(\alpha,\{\Phi_j\})$ can be bounded independent of α (cf. [17]). As mentioned above, this leads to the aim to construct coarse basis functions with minimal or bounded energy. It is also at the heart of matrix-dependent prolongation operators in multigrid methods.

For certain binary coefficient distributions, e.g. for high-permeability inclusions in a low-permeability medium as depicted in Fig. 1g, it was then possible in [8] to show (rigorously) under the assumption $\alpha \gtrsim 1$ that multiscale FEs (w.r.t. some coarse mesh $\mathscr{T}_H$) can provide such a basis $\{\Phi_j\}$, and that the indicators can be bounded independent of the contrast in α. However, they depend on H/ε, where ε is the minimum width of any island/gap.

Similarly, it was possible in [17] to show (again assuming $\alpha \gtrsim 1$) that aggregation based on a strong connection criterion (originally designed for AMG methods)

leads to a coarse basis $\{\Phi_j\}$ for which the robustness indicators can be bounded independent of the contrast in α. Here the bounds depend on H/h, since the overlap between any two supports is only $\mathscr{O}(h)$.

However, this approach to analyse robustness fails even for the simpler, reverse situation of a high-permeability medium with low-permeability inclusions (e.g. Fig. 1f), since in this case $\gamma_2(\alpha,\{\Phi_j\})$ and $\gamma_\infty(\alpha,\{\Phi_j\})$ depend on the contrast in α for any choice of $\{\Phi_j\}$. Clearly a different quasi-interpolant Π is needed in general.

3.2 Weighted Quasi-interpolant and Poincaré's Inequality

The next approach to try to prove the assumptions in Theorem 1 makes use of the weighted quasi-interpolant

$$\Pi v \;=\; \sum_{j=1}^N \bar{v}^\alpha_{\omega_j}\Phi_j, \qquad \text{where} \qquad \bar{v}^\alpha_{\omega_j} \;=\; \int_{\omega_j}\alpha v \Big/ \int_{\omega_j}\alpha\,.$$

We describe this approach for one of the simplest coarse spaces, the piecewise linear one. The following is taken from [19] (see also [6] for earlier results). Let V_H be the continuous, piecewise linear FE space associated with a shape-regular simplicial triangulation $\mathscr{T}_H$ of Ω, such that $\mathscr{T}_h$ is a refinement of $\mathscr{T}_H$. The functions $\{\Phi_j\}_{j=1}^N$ are the standard nodal basis for V_H. For simplicity, we assume that $\{\Omega_k\}_{k=1}^K = \{\omega_j\}_{j=1}^N$, and choose $\chi_k = \Phi_k$ (suitably modified near $\partial\Omega$), so that the assumptions on $\{\chi_k\}$ are satisfied with $\delta_k \sim H_k$.

The key observation in [19] is now that one further assumption suffices to fully describe the dependency of the constants C_1 and C_2 in (3) on α:

Assumption 1 Let $\omega_T = \bigcup_{\{k:\omega_k\cap T\neq\emptyset\}}\omega_k$ and $H_T = \mathrm{diam}(\omega_T)$, for $T\in\mathscr{T}_H$, and assume that there exists a $C_T^*>0$ such that, for all $v\in V_h$, either

$$\inf_{c\in\mathbb{R}}\int_{\omega_T}\alpha(v-c)^2\,d\mathbf{x} \;\lesssim\; C_T^* H_T^2\int_{\omega_T}\alpha|\nabla v|^2\,d\mathbf{x}, \qquad \text{or} \tag{4}$$

$$\partial\omega_T\cap\partial\Omega\neq\emptyset \quad\text{and}\quad \int_{\omega_T}\alpha v^2\,d\mathbf{x} \;\lesssim\; C_T^* H_T^2\int_{\omega_T}\alpha|\nabla v|^2\,d\mathbf{x}. \tag{5}$$

Proposition 1. *Let Assumption 1 hold. Then* $C_1+C_2 \lesssim C^* = \max_{T\in\mathscr{T}_H} C_T^*$.

Proof. Let $v\in V_h$ and $v_0 = \sum_{j=1}^N \bar{v}^\alpha_{\omega_j}\Phi_j$. By the Cauchy-Schwarz inequality we have $|\bar{v}^\alpha_{\omega_j}|^2 \le \int_{\omega_j}\alpha v^2 \big/ \int_{\omega_j}\alpha$, and so, using the fact that $\Phi_j\le 1$,

$$\int_T \alpha v_0^2 \;\le\; \sum_{j:\omega_j\cap T\neq\emptyset}\frac{\int_{\omega_j}\alpha v^2}{\int_{\omega_j}\alpha}\int_T\alpha\Phi_j^2 \;\le\; \int_{\omega_T}\alpha v^2,$$

which also implies $\int_T\alpha(v-v_0)^2 \lesssim \int_{\omega_T}\alpha v^2$. Now, multiplying the left hand side by $|\nabla\chi_k|_T^2$ (which is a constant $\sim H_T^{-2}$) and summing over $k\ge 1$, we get

$$\sum_{k=1}^K \|(v-v_0)\nabla\chi_k\|^2_{0,\alpha,T} \;\lesssim\; H_T^{-2}\int_{\omega_T}\alpha v^2\,. \tag{6}$$

If $\{\Phi_j\}$ forms a partition of unity on all of ω_T (i.e. if $\partial\omega_T \cap \partial\Omega = \emptyset$), we can replace v in (6) by $\hat{v} = v - c$, for any $c \in \mathbb{R}$, without changing the integral on the left hand side. Otherwise we set $\hat{v} = v$. In both cases, by Assumption 1

$$\int_{\omega_T} \alpha \hat{v}^2 \lesssim C_T^* H_T^2 \int_{\omega_T} \alpha |\nabla v|^2 . \tag{7}$$

Combining (6) and (7) and summing over all $T \in \mathscr{T}_H$ gives the bound for C_2.

The bound for C_1 can be established in a similar way (cf. [19, Lemma 4.1]).

Assumption 1 postulates the existence of a discrete weighted Poincaré/ Friedrichs–type inequality on each ω_T. It always holds, but in general the constants C_T^* will not be independent of $\alpha|_{\omega_T}$ and H_T/h. As described in detail in [19, Sect. 3] (see also [13–15]), to obtain independence of α, we require a certain local quasi–monotonicity of α on each of the regions ω_T.

Weighted Poincaré Inequalities. Let us consider a generic coarse element $T \in \mathscr{T}_H$ and define the following subsets of ω_T where α is constant:

$$\omega^m = \omega_T \cap \mathscr{Y}_m, \qquad m = 1, \ldots, M.$$

By $\mathscr{I}_T \subset \{1, \ldots, M\}$ we denote the index set of all regions ω^m that are non-empty. Let us assume w.l.o.g. that each of these subregions is connected. We generalise now the notion of quasi-monotonicity coined in [3] by considering the following three (two) directed combinatorial graphs $\Gamma^{(k)} = (\mathbf{N}, \mathscr{E}^{(k)})$, $0 \le k \le d-1$, where $\mathbf{N} = \{\omega^m : m \in \mathscr{I}_T\}$ and the edges are ordered pairs of vertices. We distinguish between three (two) different types of connections.

Definition 1. *Suppose that $\gamma^{m,m_2} = \overline{\omega}^m \cap \overline{\omega}^{m_2}$ is a non-empty manifold of dimension k, for $0 \le k \le d-1$. The ordered pair (ω^m, ω^{m_2}) is an edge in $\mathscr{E}^{(k)}$, if and only if $\alpha_m \lesssim \alpha_{m_2}$. The edges in $\mathscr{E}^{(k)}$ are said to be of type-k.*

In addition, for $1 \le k \le d-1$, we assume that

- $\mathrm{meas}(\gamma^{m,m_2}) \sim \mathrm{meas}(\omega^m \cup \omega^{m_2})^{k/d}$, and
- γ^{m,m_2} is sufficiently regular, i.e. it is a finite union of shape–regular k-dimensional simplices of diameter $\sim \mathrm{meas}(\gamma^{m,m_2})^{1/k}$.

Quasi-monotonicity is related to the connectivity in $\Gamma^{(k)}$. Let $m_* \in \mathscr{I}_T$ be the index of the region ω^{m_*} with the largest coefficient: $\alpha_{m_*} = \max_{m \in \mathscr{I}_T} \alpha_m$.

Definition 2. *The coefficient α is* type-k quasi-monotone *on ω_T, if there is a path in $\Gamma^{(k)}$ from any vertex ω^m to ω^{m_*}.*

The following lemma summarises the results in [13–15]. The existence of a benign constant C_T^* that is independent of α is directly linked to quasi-monotonicity, the way in which C_T^* depends on H_T/h to the type.

Lemma 1. *Let $\omega_T \subset \mathbb{R}^d$, $d = 2, 3$. If α is type-k quasi-monotone on ω_T, then (4) holds with*

$$C_T^* = \begin{cases} 1, & \text{if } k = d-1, \\ 1 + \log\left(\frac{H_T}{h}\right), & \text{if } k = d-2, \\ \frac{H_T}{h}, & \text{if } k = d-3. \end{cases} \tag{8}$$

A similar result can also be established in the case where $\partial\omega_K \cap \partial\Omega \neq \emptyset$, i.e. the case of Friedrichs inequality (5), see e.g. [19, Sect. 3] for details.

Quasi–monotonicity is crucial. If the coefficient is not quasi-monotone, e.g. the situation in Fig. 1d, then C^* cannot be bounded independent of α. See [19, Example 3.1] for a counter example. If the coarse mesh is not adjusted in certain critical areas of Ω, then V_H is in general not robust. The numerical results in [19] show that this is indeed the case and that quasi–monotonicity is necessary and sufficient. However, a few simple adjustments suffice, namely $\mathscr{T}_H$ has to be sufficiently fine in certain "critical" areas of Ω:

1. Choose $H_T \leq \varepsilon_m$, for all $T \in \mathscr{T}_H$ that intersect a region $\mathscr{Y}_m$ that is bordered by two regions $\mathscr{Y}_{m'}$ and $\mathscr{Y}_{m''}$ with $\alpha_{m'} \gg \alpha_m$ and $\alpha_{m''} \gg \alpha_m$. Here ε_m denotes the width of $\mathscr{Y}_m$ at its narrowest point. This ensures that α is quasi-monotone on all regions ω_T that intersect $\mathscr{Y}_m$.
2. Choose $H_T \lesssim h$, near any point or edge where α is only type-$(d-2)$ or type-$(d-3)$ quasi–monotone, i.e. near any cross point.

Usually a logarithmic growth $C^* \sim \max_T \log(H_T/h)$ is acceptable, and so even regions where the coefficient is type-$(d-2)$ quasi-monotone do not require any particular attention.

For an arbitrary piecewise constant coefficient function α there will often only be a relatively small (fixed) number of regions ω_T where α is not quasi-monotone (see e.g. Fig. 1b, e). Therefore it is very easy to ensure through some local refinement of $\mathscr{T}_H$ near these regions that $C^* \sim 1$ (or $C^* \sim \log(H/h)$). Note that crucially, this local refinement does not mean that $\mathscr{T}_H$ has to be aligned with coefficient jumps anywhere in Ω. The coarse grid merely has to be sufficiently fine in regions where α is not quasi-monotone. Ideas on how to adapt $\mathscr{T}_H$ in such a way are suggested in [19].

"Exotic" coarse spaces. Substructuring–type ("exotic") coarse spaces (as suggested in [3, 4, 16]) can be analysed in a similar way. Here the coarse basis functions are constructed as a-harmonic extensions of face, edge or vertex "cut" functions associated with a non-overlapping decomposition $\mathscr{T}_H$ of the domain. This decomposition may be related to the overlapping partitioning $\{\Omega_k\}$, or it may come from a separate coarse grid (not necessarily simplicial). If the coefficient does not vary along any of the edges/faces of $\mathscr{T}_H$, then the space can be analysed like the piecewise linear one above, using in addition the energy minimising property of the a-harmonic extension (cf. [13]). If the coefficient does vary along an edge/face, then special weighted Poincaré inequalities for functions with vanishing weighted averages across edges/-faces are required. These have recently been introduced in the context of FETI-DP methods in [12], which also analyses the robustness of the "cut" functions. An explicit analysis in the context of overlapping Schwarz does not yet exist.

3.3 Abstract Minimisation with Functional Constraints

An alternative to refining the coarse mesh in regions where α is not type–$(d-1)$ or type–$(d-2)$ quasi-monotone, is to associate more than one basis function (with possibly identical supports) with each subdomain Ω_k. Let

$$V_0 = \operatorname{span}\{\Phi_{k,j} = I_h\left(\chi_k \Psi_{k,j}\right) : j = 1,\dots,N_k,\ \ k = 1,\dots,K\},$$

where $\Psi_{k,j}$, $j = 1,\dots,N_k$, are suitable FE functions in $V_h(\overline{\Omega}_k)$ (that do not vanish on $\partial\Omega_k$) such that the functions $\{\Phi_{k,j}\} \subset V_h$ are linearly independent. Good choices for the functions $\Psi_{k,j}$ are the lowest modes of local eigenproblems, or more generally, energy minimising functions that satisfy suitable constraints. The following analysis is from [18] (see [2, 7] for related work).

In particular, let us assume that, for every Ω_k, we have a collection of linear functionals $\{f_{k,j}\}_{j=1}^{N_k} \subset V_h(\overline{\Omega}_k)'$ and let

$$\Psi_{k,j} = \arg\min_{v \in V_h(\overline{\Omega}_k)} |v|_a^2, \quad \text{subject to} \quad f_{k,l}(\Psi_{k,j}) = \delta_{jl} \quad j,l = 1,\dots,N_k\,. \tag{9}$$

Now, for any $v \in V_h$, choose the following quasi-interpolant

$$\Pi v = \sum_{k=1}^{K} I_h\left(\chi_k \Pi_{\Omega_k} v\right), \quad \text{where} \quad \Pi_{\Omega_k} v = \sum_{j=1}^{N_k} f_{k,j}(v|_{\Omega_k})\Psi_{k,j},$$

i.e. a linear combination of the basis functions $\Phi_{k,j}$ with weights $f_{k,j}(v|_{\Omega_k})$. Then the bounds on C_1 and C_2 in Theorem 3 depend only on the stability and on the local L_2-approximation properties of Π_{Ω_k} on each Ω_k.

Theorem 1. *For all $k = 1,\dots,K$ and for all $v \in V_h(\overline{\Omega}_k)$, let*

$$\|\Pi_{\Omega_k} v\|_{a,\Omega_k}^2 \le \|v\|_{a,\Omega_k}^2 \ \text{ and } \ \|v - \Pi_{\Omega_k} v\|_{0,\alpha,\Omega_k}^2 \lesssim \operatorname{diam}(\Omega_k)^2 \|u\|_{a,\Omega_k}^2\,. \tag{10}$$

Then $C_1 = \mathcal{O}(1)$ and $C_2 \lesssim (\operatorname{diam}(\Omega_k)/\delta_k)^2$.

Proof. See [18, Theorem 5.1].

Note that the minimisation problems in (9) are local to each subdomain. There are suitable choices for the functionals $f_{k,j}$ that guarantee (10) and that lead to practical algorithms to construct the functions $\Psi_{k,j}$, $j = 1,\dots,N_k$:

- $f_{k,j}(v) = (\Psi_{k,j}, v)_{0,\alpha,\Omega_k}$ where $\Psi_{k,j}$ is the jth eigenfunction corresponding to the variational eigenproblem: Find $\eta \in V_h(\overline{\Omega}_k)$ and $\lambda \ge 0$, such that

$$a(\eta, w) = \lambda(\eta, w)_{0,\alpha,\Omega_k}, \quad \text{for all} \quad w \in V_h(\overline{\Omega}_k). \tag{11}$$

  This has first been suggested and analysed in [7].
- $f_{k,j}(v) = (\Psi_{k,j}, v)_{0,\alpha,\partial\Omega_k}$ where $\Psi_{k,j}$ is the jth eigenfunction corresponding to a variational eigenproblem similar to (11), but with $(\eta, w)_{0,\alpha,\partial\Omega_k}$ instead of $(\eta, w)_{0,\alpha,\Omega_k}$ on the right hand side of (11), i.e. an eigenproblem of Steklov-Poincaré type. This has been analysed in [2].
- $f_{k,j}(v) = \overline{v}_{D_{k,j}}^{\alpha}$ where $\{D_{k,j}\}_{j=1}^{N_k}$ is a suitable non-overlapping partitioning of Ω_k such that the weighted Poincaré inequality (4) holds on each $D_{k,j}$ (e.g. $D_{k,j} = \Omega_k \cap \mathscr{Y}_j$). The construction of $\{\Psi_{k,j}\}$ requires the solution of N_k local saddle point systems and was suggested and analysed in [18].

It has been shown in [2, 7] how (10) can be proved (directly) in the first two cases, essentially based on the observation that the coarse space consists of the lowest modes corresponding to the operator pencil associated to the energy and to the weighted L_2-norm. But the assumptions can be proved for a much wider class of functionals using the following abstract approximation result in [18]. This result is related to the classical Bramble-Hilbert lemma.

Abstract Approximation Result. Consider an abstract symmetric and continuous bilinear form $a(\cdot,\cdot) : V \times V \mapsto \mathbb{R}$, as well as a collection of linear functionals $\{f_l\}_{l=1}^m \subset V'$, where $V \subset \mathscr{H}$ and $\mathscr{H}$ is a Hilbert space with norm $\|\cdot\|$. We make the following assumptions on $a(\cdot,\cdot)$, V, $\mathscr{H}$, $\|\cdot\|$ and $\{f_l\}$:

A1. $a(\cdot,\cdot)$ is positive semi-definite and defines a semi-norm $|\cdot|_a$ on V, i.e.

$$|v|_a^2 = a(v,v) \geq 0, \quad \text{for all} \quad v \in V.$$

In addition, for $v \in V$, the expression $\sqrt{\|v\|^2 + |v|_a^2}$ defines a norm on V.

A2. Let c_q be a generic constant. For all $\mathbf{q} \in \mathbb{R}^m$ there exists a $v_\mathbf{q} \in V$ with

$$f_l(v_\mathbf{q}) = q_l, \quad \text{and} \quad \|v_\mathbf{q}\| \lesssim c_q \|\mathbf{q}\|_{l^2(\mathbb{R}^m)}.$$

A3. There are two constants c_a and c_f such that

$$\|v\|^2 \leq c_a |v|_a^2 + c_f \sum_{l=1}^m |f_l(v)|^2, \quad \text{for all } v \in V. \tag{12}$$

Now, as in the specific case above, define for all $v \in V$,

$$\pi v = \sum_{l=1}^m f_l(v)\psi_l, \quad \text{where} \quad \psi_l = \arg\min_{v \in V} |v|_a^2, \quad \text{subject to} \quad f_l(\psi_j) = \delta_{jl}.$$

Then the following inequalities hold; see [18, Theorem 3.3].

Theorem 3. *Let Assumptions* **A1–A3** *be satisfied. Then, for all $u \in V$:*

$$|\pi u|_a \leq |u|_a \qquad \textit{and} \qquad \|u - \pi u\| \leq \sqrt{c_a}|u|_a. \tag{13}$$

(Note that they are independent of the constants c_q and c_f in **A2** *and* **A3**.*)*

In the specific case considered above, on an arbitrary subdomain Ω_k, Assumption **A1** is naturally satisfied with $\mathscr{H} = L_2(\Omega_k)$ and $\|\cdot\| = \|\cdot\|_{0,\alpha,\Omega_k}$. Assumption **A2** merely ensures that the linear functionals are linearly independent. Thus, the question of coarse space robustness is reduced to verifying Assumption **A3**. For one functional, i.e. for $m = 1$, this reduces to the weighted Poincaré inequality in Sect. 3.2 and to the restrictions on the coefficients made there. For more than one functional, it opens the possibility to get coefficient robustness even in the case of non-quasi-monotone coefficients, such as those depicted in Fig. 1b, d and even h. See [2, 7, 18] for the complete analysis and some numerical experiments that confirm the robustness for the functionals defined on the previous page. See also [20] for a more recent extension to systems of elliptic PDEs (such as linear elasticity).

Bibliography

[1] J.H. Bramble and J. Xu. Some estimates for a weighted L2 projection. *Math. Comput.*, 56:436–476, 1991.

[2] V. Dolean, F. Nataf, R. Scheichl, and N. Spillane. Analysis of a Two-level Schwarz Method with Coarse Spaces Based on Local Dirichlet-to-Neumann Maps. *Comput. Methods Appl. Math.*, 12(4):391–414, 2012.

[3] M. Dryja, M. Sarkis, and O.B. Widlund. Multilevel Schwarz methods for elliptic problems with discontinuous coefficients in three dimensions. *Numer. Math.*, 72:313–348, 1996.

[4] M. Dryja, B.F. Smith, and O.B. Widlund. Schwarz analysis of iterative substructuring algorithms for elliptic problems in three dimensions. *SIAM J Numer. Anal.*, 31:1662–1694, 1994.

[5] Y. Efendiev, J. Galvis, R. Lazarov, and J. Willems. Robust domain decomposition preconditioners for abstract symmetric positive definite bilinear forms. *ESAIM: Math. Model. Numer. Anal.*, 46(5):1175–1199, 2012.

[6] J. Galvis and Y. Efendiev. Domain decomposition preconditioners for multiscale flows in high contrast media. *Multiscale Model. Sim.*, 8(4):1461–1483, 2010.

[7] J. Galvis and Y. Efendiev. Domain decomposition preconditioners for multiscale flows in high contrast media: reduced dimension coarse spaces. *Multiscale Model. Sim.*, 8(5):1621–1644, 2010.

[8] I.G. Graham, P. Lechner, and R. Scheichl. Domain decomposition for multiscale PDEs. *Numer. Math.*, 106:589–626, 2007.

[9] I.G. Graham and R. Scheichl. Robust domain decomposition algorithms for multiscale PDEs. *Numer. Meth. Part. D. E.*, 23:859–878, 2007.

[10] J. Mandel and M. Brezina. Balancing domain decomposition for problems with large jumps in coefficients. *Math. Comp.*, 65:1387–1401, 1996.

[11] J. Mandel, M. Brezina, and P. Vanek. Energy optimisation of algebraic multigrid bases. *Computing*, 62:205–228, 1999.

[12] C. Pechstein, M. Sarkis, and R. Scheichl. New Theoretical Coefficient Robustness Results for FETI-DP. In *Proceedings of the 20th International Conference on Domain Decomposition Methods* (DD20), San Diego, Feb 2011, Springer, 2013.

[13] C. Pechstein and R. Scheichl. Analysis of FETI methods for multiscale PDEs – part II: Interface variation. *Numer. Math.*, 118(3):485–529, 2011.

[14] C. Pechstein and R. Scheichl. Weighted Poincaré inequalities and applications in domain decomposition. In Y. Huang et al., editors, *Domain Decomposition Methods in Science and Engineering XIX*, volume 78 of *Lecture Notes in Computational Science and Engineering*, 2011.

[15] C. Pechstein and R. Scheichl. Weighted Poincare Inequalities. *IMA J. Numer. Anal.*, Published online October 15, 2012, doi:10.1093/ imanum/drs017.

[16] M. Sarkis. Nonstandard coarse spaces and Schwarz methods for elliptic problems with discontinuous coefficients using non-conforming elements. *Numer. Math.*, 77:383–406, 1997.

[17] R. Scheichl and E. Vainikko. Additive Schwarz and aggregation-based coarsening for elliptic problems with highly variable coefficients. *Computing*, 80:319–343, 2007.

[18] R. Scheichl, P.S. Vassilevski, and L. Zikatanov. Weak approximation properties of elliptic projections with functional constraints. *Multiscale Model. Simul.*, 9(4):1677–1699, 2011.

[19] R. Scheichl, P.S. Vassilevski, and L.T. Zikatanov. Mutilevel Methods for Elliptic Problems with Highly Varying Coefficients on Non-Aligned Coarse Grids. *SIAM J. Numer. Anal.*, 50(3):1675–1694, 2012.

[20] N. Spillane, V. Dolean, P. Hauret, F. Nataf, C. Pechstein, and R. Scheichl. Abstract robust coarse spaces for systems of PDEs via generalized eigenproblems in the overlaps. NuMa-Report 2011–07, Inst. Comput. Mathematics, J.K. University Linz, 2011. Submitted.

[21] A. Toselli and O. Widlund. *Domain Decomposition Methods – Algorithms and Theory*. Springer, Berlin, 2005.

[22] P.S. Vassilevski. *Multilevel Block-Factorization Preconditioners: Matrix-based Analysis and Algorithms for Solving FE Equations*. Springer, 2008.

[23] J. Xu. Iterative methods by space decomposition and subspace correction. *SIAM Review*, 34:581–613, 1992.

[24] J. Xu and Y. Zhu. Uniform convergent multigrid methods for elliptic problems with strongly discontinuous coefficients. *Math. Mod. Meth. Appl. S.*, 18:1–29, 2008.

[25] J. Xu and J. Zou. Some nonoverlapping domain decomposition methods. *SIAM Review*, 40:857–914, 1998.

Multi-level Decompositions of Electronic Wave Functions

Harry Yserentant

Institut für Mathematik, Technische Universität Berlin, 10623 Berlin, Germany
yserentant@math.tu-berlin.de

1 Introduction

The approximation of high-dimensional functions, whether they be given explicitly or implicitly as solutions of differential equations, represents one of the grand challenges of applied mathematics. High-dimensional problems arise in many fields of application such as data analysis and statistics, but first of all in the sciences. One of the most notorious and complicated problems of this type is the Schrödinger equation. The Schrödinger equation forms the basis of quantum mechanics and is of fundamental importance for our understanding of atoms and molecules. It links chemistry to physics and describes a system of electrons and nuclei that interact by Coulomb attraction and repulsion forces. As proposed by Born and Oppenheimer in the nascency of quantum mechanics, the slower motion of the nuclei is mostly separated from that of the electrons. This results in the electronic Schrödinger equation, the problem to find the eigenvalues and eigenfunctions of the Hamilton operator

$$H = -\frac{1}{2}\sum_{i=1}^{N}\Delta_i - \sum_{i=1}^{N}\sum_{\nu=1}^{K}\frac{Z_\nu}{|\mathbf{x}_i-\mathbf{a}_\nu|} + \frac{1}{2}\sum_{\substack{i,j=1\\ i\neq j}}^{N}\frac{1}{|\mathbf{x}_i-\mathbf{x}_j|}. \tag{1}$$

It acts on functions with arguments $\mathbf{x}_1,\dots,\mathbf{x}_N \in \mathbb{R}^3$, which are associated with the positions of the considered electrons. The $\mathbf{a}_\nu$ are the fixed positions of the nuclei and the values Z_ν the charges of the nuclei in multiples of the absolute electron charge.

The high dimensionality of the equation immediately rules out classical discretization methods for partial differential equations as numerical analysts are familiar with. To overcome this curse of dimensionality, procedures like the Hartree-Fock method and its many variants and successors or density functional theory based methods have been developed over the decades. They are used with much success and form the basis of a steadily expanding branch of chemistry. See [6] for an overview on the present state of the art in quantum chemistry, and [3, 10], and [11] for mathematically oriented expositions. All these methods suffer, however, either from a priori modeling errors or from the fact that it is not clear how the accuracy can be systematically improved without the effort truly exploding for larger numbers of electrons.

R. Bank et al. (eds.), *Domain Decomposition Methods in Science and Engineering XX*, Lecture Notes in Computational Science and Engineering 91,
DOI 10.1007/978-3-642-35275-1_6,

It is therefore rather surprising that simple sparse grid-like multi-level expansions of the electronic wave functions can be constructed whose convergence rate, measured in terms of the number of basis functions involved, is independent of the number of electrons and does not much differ from that for a two- or even one-electron system. The purpose of this note is to explain these results and the effects behind them. For details we refer to the references.

2 Regularity and Decay of the Wave Functions

The at least asymptotically, in relation to the high space dimension rapid convergence of these expansions is based on very particular properties of the solutions of the electronic Schrödinger equation: their regularity, that surprisingly increases with the number of electrons, the decay behavior of their mixed derivatives, and their antisymmetry enforced by the Pauli principle.

The solution space of the electronic Schrödinger equation is first the Hilbert space H^1 that consists of the square integrable functions

$$u : (\mathbb{R}^3)^N \to \mathbb{R} : (\mathbf{x}_1, \ldots, \mathbf{x}_N) \to u(\mathbf{x}_1, \ldots, \mathbf{x}_N) \tag{2}$$

with square integrable first-order weak derivatives; the dimension of their domain increases with the number N of electrons. The norm $\|\cdot\|_1$ on H^1 is composed of the L_2-norm $\|\cdot\|_0$ induced by the L_2-inner product and the L_2-norm of the gradient. In the language of physics, the space H^1 is the space of the wave functions for which the total position probability remains finite and the expectation value of the kinetic energy can be given a meaning. It can be shown that the second-order differential operator (1) induces a bounded bilinear form on H^1 that satisfies a Garding inequality. The mathematically precise formulation of the eigenvalue problem is therefore the corresponding weak form of the equation on the space H^1, the same kind of weak form that one knows from the finite element method. The physically admissible solutions are components $u(\mathbf{x}) = \psi(\mathbf{x}, \boldsymbol{\sigma})$ of a full, spin-dependent wave function. By the Pauli principle, they are therefore antisymmetric with respect to the exchange of the positions $\mathbf{x}_i$ of electrons of the same spin $\sigma_i = \pm 1/2$.

To describe the regularity properties of the eigenfunctions, we need to introduce a scale of norms that are defined in terms of Fourier transforms. We first introduce the polynomials

$$P_{\mathrm{iso}}(\boldsymbol{\omega}) = 1 + \sum_{i=1}^{N} |\boldsymbol{\omega}_i|^2, \quad P_{\mathrm{mix}}(\boldsymbol{\omega}) = \prod_{i=1}^{N} \big(1 + |\boldsymbol{\omega}_i|^2\big). \tag{3}$$

The $\boldsymbol{\omega}_i \in \mathbb{R}^3$ forming together the variable $\boldsymbol{\omega} \in (\mathbb{R}^3)^N$ can be associated with the momentums of the electrons. The expressions $|\boldsymbol{\omega}_i|$ are their euclidean norms. The norms describing the smoothness of the solutions are now given by

$$|||u|||_{\vartheta,m}^2 = \int P_{\mathrm{iso}}(\boldsymbol{\omega})^m P_{\mathrm{mix}}(\boldsymbol{\omega})^{\vartheta} \, |\widehat{u}(\boldsymbol{\omega})|^2 \, \mathrm{d}\boldsymbol{\omega}. \tag{4}$$

They are defined on the Hilbert spaces $H^{\vartheta,m}_{\mathrm{mix}}$ that consist of the square integrable functions (2) for which these expressions remain finite. For nonnegative integer values m and ϑ, the norms measure the L_2-norm of weak partial derivatives. The parameter m measures the isotropic smoothness that does not distinguish between different directions, and the parameter ϑ the mixed smoothness in direction of the three-dimensional coordinate spaces of the electrons. The spaces L_2 and H^1 are special cases of such spaces.

It has been proved in [12] and [13] that the physically admissible eigenfunctions u of the electronic Schrödinger operator (1) are at least contained in $H^{\vartheta,1}_{\mathrm{mix}}$ for $\vartheta = 1/2$. Recently we were able to improve this result substantially. We have shown in [9] that the eigenfunctions u of the electronic Schrödinger operator are, independent of their symmetry properties, contained in

$$H^{1,0}_{\mathrm{mix}} \cap \bigcap_{\vartheta<3/4} H^{\vartheta,1}_{\mathrm{mix}}. \tag{5}$$

The bound $3/4$ is optimal and can, except for special cases, neither be reached nor improved further. The proof is based on a representation of the eigenfunctions that has been derived in [15] and for the two-electron case in [1]. It has been shown in [15] that the eigenfunctions can be written as products

$$u(\mathbf{x}) = \exp\left(\sum_{i<j} \phi(\mathbf{x}_i - \mathbf{x}_j)\right) v(\mathbf{x}) \tag{6}$$

of more regular functions $v \in H^{1,1}_{\mathrm{mix}}$ and a universal factor that covers their singularities. This kind of splitting can be traced back to the work of Hylleraas [8] in the early years of quantum mechanics. It has been used in [4] and [7] to study the Hölder regularity of the eigenfunctions. There is a lot of freedom in the choice of the function ϕ. It needs only to be of the form

$$\phi(\mathbf{x}) = \widetilde{\phi}(|\mathbf{x}|), \quad \widetilde{\phi}'(0) = \frac{1}{2}, \tag{7}$$

where $\widetilde{\phi} : [0,\infty) \to \mathbb{R}$ is an infinitely differentiable function behaving sufficiently well at infinity. The regularity is therefore determined by that of the explicitly known factor from (6) that describes the behavior of the solutions at the singular points of the electron-electron interaction potential.

The splitting (6) is of independent interest since it is obviously possible to obtain better convergence rates for the regular part of the solutions than for the solutions themselves. We will restrict ourselves, however, here to the direct approximation of the eigenfunctions. The domain of the eigenfunctions is infinitely extended. The eigenfunctions are, however, strongly localized. It is known for a long time that an eigenfunction u for an eigenvalue below the ionization threshold of the given atom or molecule decays exponentially in the L_2-sense. That means there is a constant $\gamma > 0$ such that the function

$$\mathbf{x} \to \exp\left(\gamma \sum_{i=1}^{N} |\mathbf{x}_i|\right) u(\mathbf{x}), \tag{8}$$

is square integrable. This constant depends on the distance of the eigenvalue under consideration to the bottom of the essential spectrum. More details and references to the literature can be found in [14]. It has been shown in [15] that these exponentially weighted eigenfunctions admit the same kind of representation (6) as the eigenfunctions themselves. Thus they share with them the described regularity properties [9]. The convergence analysis is based on this observation.

3 Sparse Grids and Antisymmetry

To explain the meaning of these results for the approximation of the solutions of the Schrödinger equation, we consider a simple model problem, the approximation of functions u of the variables $x_1,\dots,x_d$ that are odd and 2π-periodic in every coordinate direction on the cube $Q=[0,\pi]^d$ by tensor products

$$\phi(\mathbf{k},\mathbf{x}) = \prod_{i=1}^{d} \phi_{k_i}(x_i) \tag{9}$$

of the one-dimensional trigonometric polynomials

$$\phi_{k_i}(\xi) = \sqrt{\frac{2}{\pi}}\,\sin(k_i\xi) \tag{10}$$

labeled by the components $k_i = 1,2,\dots$ of the multi-indices $\mathbf{k}$. Our presentation closely follows [14]. Functions of the given kind that are square integrable over Q can be expanded into a multivariate Fourier series

$$u(\mathbf{x}) = \sum_{\mathbf{k}} \widehat{u}(\mathbf{k})\phi(\mathbf{k},\mathbf{x}), \tag{11}$$

where the expansion coefficients are given by

$$\widehat{u}(\mathbf{k}) = \int_Q u(\mathbf{x})\phi(\mathbf{k},\mathbf{x})\,\mathrm{d}\mathbf{x}. \tag{12}$$

We measure the speed of convergence of this series in the sense of the L_2-norm which reads in terms of the expansion coefficients

$$\|u\|_0^2 = \sum_{\mathbf{k}} |\widehat{u}(\mathbf{k})|^2. \tag{13}$$

The speed of convergence of the series is therefore determined by the speed with which the expansion coefficients decay. Assume that all partial derivatives of u of order s exist and are square integrable. This implies that

$$|u|_s^2 = \sum_{\mathbf{k}} |\mathbf{k}|^{2s} |\widehat{u}(\mathbf{k})|^2 \tag{14}$$

remains finite, where $|\mathbf{k}|$ is defined by

$$|\mathbf{k}|^2 = \sum_{i=1}^{d} k_i^2. \tag{15}$$

Consider now the finite part u_ε of the series (11) that extends over the multi-indices $\mathbf{k}$ inside the ball of radius $1/\varepsilon$ around the origin, for which

$$|\mathbf{k}| < \frac{1}{\varepsilon}. \tag{16}$$

Due to the orthonormality of the functions (9), u_ε is the best approximation of u by a linear combination of the selected basis functions. It holds

$$\|u - u_\varepsilon\|_0^2 \leq \varepsilon^{2s} \sum_{\mathbf{k}} |\mathbf{k}|^{2s} |\widehat{u}(\mathbf{k})|^2 = \varepsilon^{2s} |u|_s^2. \tag{17}$$

The number n of these basis functions grows like

$$n \sim \frac{1}{\varepsilon^d} \tag{18}$$

as ε goes to zero. This is out of every reach for higher space dimensions d, the curse of dimensionality. It can only be broken if one restricts oneself to a class of functions whose smoothness increases sufficiently fast with the space dimension d. At this place the mixed regularity comes into play. Consider functions u that possess corresponding weak partial derivatives and set

$$|u|_{1,\text{mix}}^2 = \int_Q \left| \frac{\partial^d u}{\partial x_1 \dots \partial x_d} \right|^2 \mathrm{d}\mathbf{x} \tag{19}$$

or, in terms of the expansion coefficients,

$$|u|_{1,\text{mix}}^2 = \sum_{\mathbf{k}} \left(\prod_{i=1}^{d} k_i \right)^2 |\widehat{u}(\mathbf{k})|^2. \tag{20}$$

Let u_ε^* be the function represented by the finite part of the series (11) that extends over the multi-indices $\mathbf{k}$ inside the hyperboloid given by

$$\prod_{i=1}^{d} k_i < \frac{1}{\varepsilon}, \tag{21}$$

instead of the ball (16). The L_2-error can then be estimated as

$$\|u - u_\varepsilon^*\|_0 \leq \varepsilon\, |u|_{1,\text{mix}} \tag{22}$$

and tends like $\mathcal{O}(\varepsilon)$ to zero. The dimension n of the space spanned by the functions (9) for which (21) holds, now increases, however, only like

$$n \sim |\log \varepsilon|^{d-1} \varepsilon^{-1}. \tag{23}$$

This shows that a comparatively slow growth of the smoothness can help to reduce the complexity substantially, an observation that forms the basis of the sparse grid or hyperbolic cross techniques; see [2] for an overview. Due to the presence of the logarithmic term, the applicability of such methods is, however, still limited to moderate space dimensions.

The rescue comes from the symmetry properties of the wave functions enforced by the Pauli principle. They represent a possibility to escape from this dilemma without forcing up the smoothness requirements further, which has first been noted by Hackbusch [5]. Consider functions u that are antisymmetric with respect to the exchange of their variables, i.e., that

$$u(\mathbf{P}\mathbf{x}) = \operatorname{sign}(\mathbf{P})u(\mathbf{x}) \tag{24}$$

holds for all permutation matrices $\mathbf{P}$. It is not astonishing that such symmetry properties are immediately reflected in the expansion (11). Let

$$\widetilde{\phi}(\mathbf{k},\mathbf{x}) = \frac{1}{\sqrt{d!}} \sum_{\mathbf{P}} \operatorname{sign}(\mathbf{P})\phi(\mathbf{k},\mathbf{P}\mathbf{x}) \tag{25}$$

be the renormalized, antisymmetric parts of the functions (9), where the sums extend over the $d!$ permutation matrices $\mathbf{P}$ of order d. The antisymmetrized functions (25) can be written as determinants

$$\frac{1}{\sqrt{d!}} \begin{vmatrix} \phi_{k_1}(x_1) & \dots & \phi_{k_d}(x_1) \\ \vdots & \ddots & \vdots \\ \phi_{k_1}(x_d) & \dots & \phi_{k_d}(x_d) \end{vmatrix} \tag{26}$$

and evaluated in this way. For the functions u in the given symmetry class, many terms in the expansion (11) can be combined. It finally collapses into

$$u(\mathbf{x}) = \sum_{k_1>\ldots>k_d} \big(u, \widetilde{\phi}(\mathbf{k},\cdot)\big)\widetilde{\phi}(\mathbf{k},\mathbf{x}), \tag{27}$$

where the expansion coefficients are the L_2-inner products of u with the corresponding functions (25). The number of basis functions needed to reach a given accuracy is reduced by more than the factor $d!$, a very significant gain for larger dimensions d.

It remains to count the number of the sequences $k_1 > k_2 > \ldots > k_d$ of natural numbers that satisfy the condition (21) and with that also the number of basis function (25) needed to reach the accuracy $\mathcal{O}(\varepsilon)$. To study the asymptotic behavior of the number of these sequences in dependence of the dimension d and the accuracy ε, it suffices when we restrict ourselves to the case $\varepsilon = 1/2^L$, with positive integers L. That is, we have to give bounds for the number of sequences $k_1 > \ldots > k_d$ for which

$$\prod_{i=1}^{d} k_i \le 2^L. \tag{28}$$

The problem to estimate this number has to do with the prime factorization of integers. To simplify this problem, we group the numbers k_i into levels and decompose the space of the trigonometric polynomials correspondingly. Let

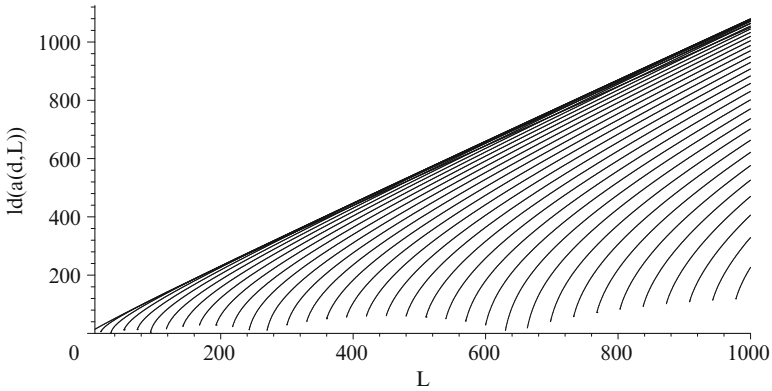


Fig. 1. The numbers $a^*(L)$ and $a(d,L)$ for $d = 10, 15, 20, \ldots, 175$

$$\ell(k_i) = \max\left\{ \ell \in \mathbb{Z} \,\middle|\, 2^{\ell} \le k_i \right\}. \tag{29}$$

An upper bound for the number of these sequences is then the number $a(d,L)$ of the sequences $k_1 > k_2 > \ldots > k_d$ of natural numbers for which

$$\prod_{i=1}^{d} 2^{\ell(k_i)} \le 2^L. \tag{30}$$

The numbers $a(d,L)$ can be calculated recursively; see [14] for details. A crude estimate yields $a(d,L) = 0$ if $L+1 < d$. Thus

$$a^*(L) := \max_{d \ge 1} a(d,L) = \max_{d \le L+1} a(d,L). \tag{31}$$

Figure 1 shows, in logarithmic scale, how the $a(d,L)$ behave compared to their joint least upper bound $a^*(L)$. It becomes obvious from this picture that this upper bound exceeds the actual dimensions for larger d by many orders of magnitude, the more the more the number d of variables increases. The joint least upper bound that is independent of d for the number of the sequences $k_1 > \ldots > k_d$ of natural numbers k_i for which (28) holds grows at least like $\sim 2^L$ since already for the case $d = 1$, there are 2^L such "sequences", namely those with values $k_1 = 1, \ldots, 2^L$. Figure 1 suggests conversely that the upper bound (31) for the number of these sequences does not grow much faster than $\sim 2^L$. This is in fact the case since the number of the decreasing infinite sequences $k_1 \ge k_2 \ge k_3 \ge \ldots$ of natural numbers for which

$$\prod_{i=1}^{\infty} 2^{\ell(k_i)} \le 2^L, \tag{32}$$

with L a given nonnegative integer, is bounded by

$$\sum_{\ell=0}^{L} p(\ell)\, 2^{\ell}, \tag{33}$$

where $p(\ell)$ denotes the partition number of ℓ, the number of possibilities of representing ℓ as sum of nonnegative integers without regard to the order. To show this, we observe that the number of these sequences is bounded by the number of sequences $k_1, k_2, k_3, \dots$ of natural numbers for which at least their levels $\ell(k_1), \ell(k_2), \dots$ decrease and that satisfy (32). We show that the expression (33) counts the number of these sequences. Let the integers $\ell_i = \ell(k_i)$ first be given. As there are 2^{ℓ_i} natural numbers k_i for which $\ell(k_i) = \ell_i$, namely $k_i = 2^{\ell_i}, \dots, 2^{\ell_i+1} - 1$, there are

$$\prod_{i=1}^{\infty} 2^{\ell_i} = 2^{\ell}, \quad \ell = \sum_{i=1}^{\infty} \ell_i, \tag{34}$$

sequences $k_1, k_2, k_3, \dots$ for which the $\ell(k_i)$ attain the prescribed values ℓ_i. The problem thus reduces to the question how many decreasing sequences of nonnegative integers ℓ_i exist that sum up to values $\ell \le L$, i.e., for which

$$\sum_{i=1}^{\infty} \ell_i = \ell. \tag{35}$$

This number is by definition the partition number $p(\ell)$ of the nonnegative integer ℓ. Every sequence $k_1 > k_2 > \dots > k_d$ of natural numbers for which (28) holds can obviously be expanded to an infinite, decreasing sequence $k_1 \ge k_2 \ge k_3 \ge \dots$ of natural numbers that satisfies the condition (32) by setting all $k_i = 1$ for $i > d$. The sum (33) represents therefore also an upper bound for the number of these sequences.

The partition number plays a big role in combinatorics. Hardy and Ramanujan have shown that it behaves asymptotically like

$$p(\ell) \sim \frac{\exp\left(\pi\sqrt{2\ell/3}\right)}{\ell} \tag{36}$$

as ℓ goes to infinity. We conclude that the upper bound (31) for the number of determinants needed to reach an error $\le 2^{-L}|u|_{1,\mathrm{mix}}$ behaves like

$$a^*(L) = (2^L)^{1+\delta(L)}, \quad 0 \le \delta(L) \le cL^{-1/2}, \tag{37}$$

where c is a constant that depends neither on L nor on the space dimension d or the function u. Using the representation of $a^*(L)$ from (31) and the recursively calculated values $a(d,L)$, the exponents $1 + \delta(L)$ can be calculated exactly. They decay for L ranging from 10 to 1,000 monotonely from 1.406 to 1.079. For $L = 100$, $1 + \delta(L) = 1.204$. That is, the error tends faster to zero in the number n of determinants than

$$\sim \frac{1}{n^{1-\vartheta}} \tag{38}$$

for any given ϑ in the interval $0 < \vartheta < 1$. Not only does the convergence rate deteriorate neither with the dimension nor the number of variables, it behaves asymptotically almost as in the one-dimensional case. Similar results hold for partially antisymmetric functions as they occur in quantum mechanics.

4 Eigenfunction and Wavelet Expansions

The constructions sketched in the previous section transfer to the more complicated case of the expansion of the solutions of the electronic Schrödinger equation into correspondingly antisymmetrized tensor products of three-dimensional Hermite functions or other eigenfunctions of three-dimensional Schrödinger-like operators as in [14] or wavelets as in [16]. Indeed, it finally turns out that the convergence rate measured in terms of the number of basis functions involved does not deteriorate with the number of electrons and comes close to that for the two- or even one-particle case. We do not explicate the partly technical details here but explain how one can utilize the intermediate smoothness of the exponentially weighted solutions (8) to obtain optimal convergence rates.

Let e^{ψ} be exponential factor in (8). The argumentation starts from functions v whose exponentially weighted counterparts $e^{\psi}v$ are located in $H^{1,1}_{\mathrm{mix}}$, that is, have in contrast to the solutions of the Schrödinger equation full mixed regularity. The essential observation is that the norm $|||e^{\psi}v|||_{1,1}$ can be estimated by the sum of the weighted L_2-norms $\|e^{\psi}D^{\boldsymbol{\alpha}}v\|_0$ of the involved derivatives $D^{\boldsymbol{\alpha}}v$ of v and vice versa. This comes from the special structure of the function ψ. The norm $|||e^{\psi}v|||_{1,1}$ measures therefore the exponentially weighted L_2-norms of the involved derivatives of v. It is therefore reasonable to start from a sequence $T_n : H^1 \to H^1$, $n = 1,2,\ldots$, of linear approximation operators that are uniformly H^1-bounded and to require that

$$\|v - T_n v\|_1 \lesssim n^{-q} |||e^{\psi}v|||_{1,1} \tag{39}$$

for all functions $v \in H^1$ for which $e^{\psi}v \in H^{1,1}_{\mathrm{mix}}$. The constant $q > 0$ is an unspecified convergence rate also depending on what n means. These assumptions form a proper framework for sparse grid-like approximation methods as those mentioned above modeled after the example from the last section. Another example is the expansion into tensor products of three-dimensional functions with given angular parts; see [14]. The range of the T_n is in this case infinite dimensional. The exponential factor is the tribute paid to the infinite extension of the domain. The assumption (39) implies for the functions $u \in H^1$ for which $e^{\psi}u \in H^{\vartheta,1}_{\mathrm{mix}}$ for some $0 < \vartheta < 1$, the error estimate

$$\|u - T_n u\|_1 \lesssim n^{-\vartheta q} |||e^{\psi}u|||_{\vartheta,1}. \tag{40}$$

The proof utilizes that the spaces $H^{\vartheta,1}_{\mathrm{mix}}$, $0 < \vartheta < 1$, are interpolation spaces between the spaces $H^1 = H^{0,1}_{\mathrm{mix}}$ and $H^{1,1}_{\mathrm{mix}}$.

We conclude that for the case of the solutions u of the Schrödinger equation the H^1-error $\|u - T_n u\|_1$ tends faster to zero as $n^{-\vartheta q}$ for any $\vartheta < 3/4$. An estimate directly based on an estimate of their K-functional even shows that

$$\|u - T_n u\|_1 \lesssim \sqrt{\ln(n)}\, n^{-3/4q} \tag{41}$$

so that up to the logarithmic term only the factor $3/4$ gets lost compared to the case of full mixed regularity. The estimate is optimal, at least up to the logarithmic factor, and can in general not be improved further.

Bibliography

[1] M. Bachmayr. Hyperbolic wavelet discretization of the two-electron Schrödinger equation in an explicitly correlated formulation. Preprint AICES-2010/06-2, RWTH Aachen, June 2010.
[2] H.-J. Bungartz and M. Griebel. Sparse grids. *Acta Numerica*, 13:1–123, 2004.
[3] E. Cancès, C. Le Bris, and Y. Maday. *Méthodes Mathématiques en Chimie Quantique*. Springer, Berlin Heidelberg New York, 2006.
[4] S. Fournais, M. Hoffmann-Ostenhof, T. Hoffmann-Ostenhof, and T. Østergard Sørensen. Sharp regularity estimates for Coulombic many-electron wave functions. *Commun. Math. Phys.*, 255:183–227, 2005.
[5] W. Hackbusch. The efficient computation of certain determinants arising in the treatment of Schrödinger's equation. *Computing*, 67:35–56, 2000.
[6] T. Helgaker, P. Jørgensen, and J. Olsen. *Molecular Electronic Structure Theory*. John Wiley & Sons, Chichester, 2000.
[7] M. Hoffmann-Ostenhof, T. Hoffmann-Ostenhof, and T. Østergard Sørensen. Electron wavefunctions and densities for atoms. *Ann. Henri Poincaré*, 2:77–100, 2001.
[8] E.A. Hylleraas. Neue Berechnung der Energie des Heliums im Grundzustande, sowie des tiefsten Terms von Ortho-Helium. *Z. Phys.*, 54:347–366, 1929.
[9] H.-C. Kreusler and H. Yserentant. The mixed regularity of electronic wave functions in fractional order and weighted Sobolev spaces. *Numer. Math.*, 121:781–802, 2012.
[10] C. Le Bris, editor. *Handbook of Numerical Analysis, Vol. X: Computational Chemistry*. North Holland, Amsterdam, 2003.
[11] C. Le Bris. Computational chemistry from the perspective of numerical analysis. *Acta Numerica*, 14:363–444, 2005.
[12] H. Yserentant. On the regularity of the electronic Schrödinger equation in Hilbert spaces of mixed derivatives. *Numer. Math.*, 98:731–759, 2004.
[13] H. Yserentant. The hyperbolic cross space approximation of electronic wavefunctions. *Numer. Math.*, 105:659–690, 2007.
[14] H. Yserentant. *Regularity and Approximability of Electronic Wave Functions*, volume 2000 of *Lecture Notes in Mathematics*. Springer, 2010.
[15] H. Yserentant. The mixed regularity of electronic wave functions multiplied by explicit correlation factors. *ESAIM: M2AN*, 45:803–824, 2011.
[16] A. Zeiser. Wavelet approximation in weighted Sobolev spaces of mixed order with applications to the electronic Schrödinger equation. *Constr. Approx.*, 35:293–322, 2012.

A Substructuring Preconditioner for Three-Dimensional Maxwell's Equations

Qiya Hu[1], Shi Shu[2] and Jun Zou[3]

[1] LSEC, Institute of Computational Mathematics and Scientific/Engineering Computing, Academy of Mathematics and Systems Science, The Chinese Academy of Sciences, Beijing 100080, China hqy@lsec.cc.ac.cn
[2] School of Mathematics and Computational Science, Xiangtan University, Hunan, 411105, China shushi@xtu.edu.cn
[3] Department of Mathematics, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong zou@math.cuhk.edu.hk

Summary. We propose a new nonoverlapping domain decomposition preconditioner for the discrete system arising from the edge element discretization of the three-dimensional Maxwell's equations. This preconditioner uses the simplest coarse edge element space induced by the coarse triangulation. We will show that the rate of the PCG convergence with this substructuring preconditioner is quasi-optimal, and is independent of large variations of the coefficients across the local interfaces.

1 Introduction

When the time-dependent Maxwell's equations are solved numerically, we need to solve the following **curlcurl**-system at each time step [4, 6, 8, 12]:

$$\mathbf{curl}(\alpha \mathbf{curlu}) + \beta \mathbf{u} = \mathbf{f} \quad \text{in} \quad \Omega \tag{1}$$

where Ω is assumed to be an open polyhedral domain in $\mathbf{R}^3$, and the coefficients $\alpha(\mathbf{x})$ and $\beta(\mathbf{x})$ are two positive bounded functions in Ω. We shall complement the Eq. (1) with the perfect conductor condition $\mathbf{u} \times \mathbf{n} = 0$ on $\partial\Omega$, where $\mathbf{n}$ is the unit outward normal vector on $\partial\Omega$.

Edge finite element methods have been widely applied in the numerical solution of the system (1), see, for example, [5, 6, 8, 11]. Compared to the standard nodal finite element methods, the discrete systems resulting from the edge element discretization are essentially different in nature. The non-overlapping domain decomposition preconditioners have been well developed for the nodal element systems for the standard second order elliptic problems in the past two decades, and proved both numerically and theoretically to perform nearly optimally in terms of the fine mesh size and subdomain size; see, e.g., the monograph [15]. But these preconditioners,

R. Bank et al. (eds.), *Domain Decomposition Methods in Science and Engineering XX*, Lecture Notes in Computational Science and Engineering 91,
DOI 10.1007/978-3-642-35275-1_7,

or their natural generalizations turn out to perform mostly very poorly for the edge element systems for the **curlcurl**-system (1), especially in three dimensions.

A lot of important efforts have been made in the construction of effective domain decomposition methods for the system (1). A substructuring type method was analysed in [16] for two dimensions, and in [2] for three dimensions with two subdomains. In [7], a novel substructuring type method was proposed for general two-dimensional multiple subdomains with quite irregular boundaries, and it was proved to be nearly optimal in terms of a variety of mesh decompositions and distributions of physical material properties. However, it has been a challenge how to construct an efficient non-overlapping domain decomposition preconditioner for the Maxwell's equations in three dimensions with general multiple subdomains. A first important attempt to this problem was made in [9] where a wire basket type algorithm was proposed and analysed. Then a substructuring preconditioner and a dual-primal FETI algorithm were introduced and fully analysed for three dimensions in [10] and [14], respectively. These three methods have their respective advantages and disadvantages: the algorithms in [9] and [14] both involve smaller coarse solvers but they are difficult to implement; the method in [10] is easier to implement but it involves a relatively large coarse solver.

This work intends to construct a new substructuring type preconditioner for the three-dimensional **curlcurl**-system (1) for general multiple subdomains. In this preconditioner, the coarse space is chosen to be the edge element space induced by the coarse triangulation, so the resulting coarse solver is very cheap and simple to implement. It is shown that the rate of the PCG convergence with this substructuring preconditioner is quasi-optimal, and more importantly, independent of the large variations of the coefficients in the system (1) across the local interfaces.

2 Domain Decompositions and Discretizations

This section introduces the non-overlapping domain decomposition of domain Ω, the weak form of the system (1) and the edge element spaces.

2.1 Initial Domain Decomposition Based on the Distribution of the Coefficients

We assume that the entire domain Ω is decomposed into N_0 open convex polyhedral subdomains $D_1, D_2, \cdots, D_{N_0}$ such that $\bar{\Omega} = \cup_{r=1}^{N_0} \bar{D}_r$ and $\alpha(\mathbf{x})$ and $\beta(\mathbf{x})$ are positive constants on each subdomain D_r, namely for $r = 1, 2, \ldots, N_0$,

$$\alpha(\mathbf{x}) = \alpha_r, \quad \beta(\mathbf{x}) = \beta_r \quad \forall \mathbf{x} \in D_r\,.$$

Clearly such a decomposition is always possible when the domain Ω is occupied by multiple media. In fact, if for some medium we have an irregular nonconvex subregion in Ω, we can further split each nonconvex medium subregion into smaller convex subdomains. This means that our assumption does cover many practical cases, especially considering the fact that the domain Ω on which we solve the original

Maxwell system (1) by a finite element method is often obtained by approximating the original physical domain by a polyhedral domain. Note that N_0 typically is a *fixed* constant in applications, so $diam(D_r) = O(1)$.

Let F_{nm} denote the common face of two neighboring subdomains D_n and D_m, and set $D_{nm} = D_n \cup D_m \cup F_{nm}$. For simplicity of the analysis, we assume

$$\beta_r \lesssim \alpha_r \lesssim d^{-2}\alpha_r, \quad r = 1, \cdots, N_0. \tag{2}$$

2.2 Domain Decomposition

For a number $d \in (0,\ 1)$, let each polyhedron D_l be decomposed into the union of some non-overlapping tetrahedra (or hexahedra) $\{\Omega_k\}$ of size d (see [3, 15] and [18]), which results in a non-overlapping domain decomposition for Ω: $\bar{\Omega} = \bigcup_{k=1}^{N} \bar{\Omega}_k$. Naturally we further assume that $\Omega_i \cap \Omega_j = \emptyset$ when $i \neq j$; if $i \neq j$ and $\partial\Omega_i \cap \partial\Omega_j \neq \emptyset$, $\partial\Omega_i \cap \partial\Omega_j$ is a common face (or edge or vertex) of Ω_i and Ω_j. Now the subdomains $\Omega_1, \cdots, \Omega_N$ constitute our desired *coarse* triangulation $\mathscr{T}_d$ of Ω. The faces and vertices of the subdomains are always denoted by F and V, while the common (open) face of the subdomains Ω_i and Ω_j are denoted by Γ_{ij}, and the union of all such common faces by Γ, i.e., $\Gamma = \cup\bar{\Gamma}_{ij}$. Γ will be called *the interface*. By Γ_k we denote the intersection of Γ with the boundary of the subdomain Ω_k. So we have $\Gamma_k = \partial\Omega_k$ if Ω_k is an interior subdomain of Ω. We shall set $\Omega_{ij} = \Omega_i \cup \Omega_j \cup \Gamma_{ij}$.

2.3 Weak Formulation

Let $H(\mathbf{curl}; \Omega)$ be the Sobolev space consisting of all square integrable functions whose **curl**'s are also square integrable in Ω, and $H_0(\mathbf{curl}; \Omega)$ be a subspace of $H(\mathbf{curl}; \Omega)$ of all functions whose tangential components vanish on $\partial\Omega$. Then by writing the scalar product in $(L^2(\Omega))^3$ as $(\cdot, \cdot)$, we can state the variational problem for system (1) as follows:

Find $\mathbf{u} \in H_0(\mathbf{curl}; \Omega)$ such that

$$\mathscr{A}(\mathbf{u}, \mathbf{v}) = (\mathbf{f}, \mathbf{v}), \quad \forall \mathbf{v} \in H_0(\mathbf{curl}; \Omega) \tag{3}$$

where $\mathscr{A}(\cdot, \cdot)$ is a bilinear form given by

$$\mathscr{A}(\mathbf{u}, \mathbf{v}) = (\alpha\ \mathbf{curl}\ \mathbf{u}, \mathbf{curl}\ \mathbf{v}) + (\beta\mathbf{u}, \mathbf{v}), \quad \mathbf{u}, \mathbf{v} \in H(\mathbf{curl}; \Omega).$$

2.4 Fine Triangulation and Their Associated Finite Element Spaces

We further divide each Ω_k into smaller tetrahedral elements of size h so that elements from two neighboring subdomains have an intersection which is either empty or a single nodal point or an edge or a face on the interface Γ. Let $\mathscr{T}_h$ be the resulting triangulation of the domain Ω, which we assume is quasi-uniform. Then we introduce the N*é*d*é*lec edge element space of the lowest order defined on $\mathscr{T}_h$ (cf. [12] and [13]):

$$V_h(\Omega) = \Big\{ \mathbf{v} \in H_0(\mathbf{curl}; \Omega);\ \mathbf{v}\,|_K \in R(K),\ \forall K \in \mathscr{T}_h \Big\},$$

where $R(K)$ is a subset of all linear polynomials on the element K of the form:

$$R(K) = \Big\{ \mathbf{a} + \mathbf{b} \times \mathbf{x};\ \mathbf{a}, \mathbf{b} \in \mathbf{R}^3,\ \mathbf{x} \in K \Big\}.$$

In an analogous way, we can define the coarse edge element space $V_d(\Omega) \subset V_h(\Omega)$, associated with the *coarse* triangulation $\mathscr{T}_d$.

It is well-known that for any $\mathbf{v} \in V_h(\Omega)$, its tangential components are continuous on all edges of each element in the triangulation $\mathscr{T}_h$. Moreover, each edge element function $\mathbf{v}$ in $V_h(\Omega)$ is uniquely determined by its moments on each edge e of $\mathscr{T}_h$:

$$\Big\{ \lambda_e(\mathbf{v}) = \int_e \mathbf{v} \cdot \mathbf{t}_e ds;\ e \in \mathscr{E}_h \Big\},$$

where $\mathscr{E}_h$ denotes the set of the *fine* edges from the triangulation $\mathscr{T}_h$, and $\mathbf{t}_e$ denotes the unit vector on the edge e.

By $Z_h(\Omega)$ we denote the continuous piecewise linear finite element subspace of $H_0^1(\Omega)$ associated with the triangulation $\mathscr{T}_h$. Similarly, let $Z_d(\Omega)$ denote the continuous piecewise linear finite element subspace of $H_0^1(\Omega)$ associated with the triangulation $\mathscr{T}_d$.

2.5 Discrete Variational Problem

Using the edge element space $V_h(\Omega)$, the system (3) may be approximated as follows: Find $\mathbf{u}_h \in V_h(\Omega)$ such that

$$(\alpha \mathbf{curl}\, \mathbf{u}_h, \mathbf{curl}\, \mathbf{v}_h) + (\beta \mathbf{u}_h, \mathbf{v}_h) = (\mathbf{f}, \mathbf{v}_h), \quad \forall \mathbf{v}_h \in V_h(\Omega). \tag{4}$$

Define the operator $A : V_h(\Omega) \to V_h(\Omega)$ by

$$(A\mathbf{u}_h, \mathbf{v}_h) = (\alpha \mathbf{curl}\, \mathbf{u}_h, \mathbf{curl}\, \mathbf{v}_h) + (\beta \mathbf{u}_h, \mathbf{v}_h), \quad \forall \mathbf{u}_h, \mathbf{v}_h \in V_h(\Omega),$$

Then, (4) can be written in the operator form

$$A\mathbf{u}_h = \mathbf{f}_h \tag{5}$$

where $\mathbf{f}_h$ is the L^2-projection of $\mathbf{f}$ onto $V_h(\Omega)$.

3 A Nearly Optimal Preconditioner for A

3.1 Construction of the Preconditioner

We first introduce some useful sets and subspaces.

$\mathscr{E}_h$: the set of all edges from the triangulations $\mathscr{T}_h$;

$\mathscr{E}_{\Gamma,h}$: the set of edges which belong to $\mathscr{E}_h$ and have two endpoints on the interface Γ;

$\mathscr{E}_d$: the set of all (coarse) edges from the triangulations $\mathscr{T}_d$;

$\mathscr{W}_{\mathrm{E}}$: the union of all the coarse edges $\mathrm{E}' \in \mathscr{E}_d$, which have a common endpoint with the coarse edge $\mathrm{E} \in \mathscr{E}_d$. And $\mathscr{W}_{\mathrm{E}}$ is called E-*basket*.

$\mathscr{E}^b_{\mathrm{E},h}$: the set of all (fine) edges which belong to $\mathscr{E}_h$ and have at least one endpoint on $\mathscr{W}_{\mathrm{E}}$;

Let D be either a subdomain D_r or a subdomain Ω_k or a subdomain Ω_{ij} or a subdomain D_{mn}. The restrictions of $V_h(\Omega)$ (resp. $Z_h(\Omega)$) on D is denoted by $V_h(D)$ (resp. $Z_h(D)$). The following local subspaces of $V_h(D)$ will be important to our analysis:

$$V_h^0(D) = \Big\{ \mathbf{v} \in V_h(D);\ \mathbf{v} \times \mathbf{n} = 0 \ \text{ on } \ \partial D \Big\},$$

and

$$Z_h^0(D) = \Big\{ \varphi \in Z_h(\Omega);\ supp\ \varphi \subset D \Big\}.$$

We define subspaces of $V_h(\Omega)$:

$$V_h^H(\Omega) = \Big\{ \mathbf{v} \in V_h(\Omega);\ \mathbf{v} \text{ is the discrete } A\text{-extension of } \mathbf{v}|_{\partial\Omega_k} \text{ in each } \Omega_k \Big\},$$

$$V_h^H(\Omega_{ij}) = V_h^H(\Omega) \bigcap V_h^0(\Omega_{ij}),$$

and for $\mathrm{E} \in \mathscr{E}_d$,

$$V_h^{\mathrm{E}}(\Omega) = \Big\{ \mathbf{v} \in V_h^H(\Omega);\ \lambda_e(\mathbf{v}) = 0 \text{ for each } e \in \mathscr{E}_{\Gamma,h} \backslash \mathscr{E}^b_{\mathrm{E},h} \Big\}.$$

It is well known that a suitable *coarse* subspace plays a key role in the construction of an effective domain decomposition preconditioner, and it is generally rather technical and problem-dependent to choose such a *coarse* subspace. Surprisingly we are going to choose the coarse subspace to be the simplest one, namely the subspace $V_d(\Omega)$ induced by the coarse triangluation $\mathscr{T}_d$.

It is easy to see that the space $V_h(\Omega)$ has the (non-direct sum) decomposition

$$V_h(\Omega) = V_d(\Omega) + \sum_{k=1}^{N} V_h^0(\Omega_k) + \sum_{\mathrm{E}} V_h^{\mathrm{E}}(\Omega) + \sum_{\Gamma_{ij}} V_h^H(\Omega_{ij}). \tag{6}$$

Next, we define the corresponding solvers on the subspaces $V_h^0(\Omega_k)$, $V_h^{\mathrm{E}}(\Omega)$, $V_h^H(\Omega_{ij})$ and $V_d(\Omega)$.

As usual, we denote the restriction of A on $V_h^0(\Omega_k)$ by A_k, i.e.,

$$(A_k\mathbf{v}, \mathbf{u})_{\Omega_k} = (A\mathbf{v}, \mathbf{u}) = \mathscr{A}(\mathbf{v}, \mathbf{u}),\ \mathbf{v} \in V_h^0(\Omega_k),\ \forall \mathbf{u} \in V_h^0(\Omega_k).$$

Let $B_k : V_h^0(\Omega_k) \to V_h^0(\Omega_k)$, $B_d : V_d(\Omega) \to V_d(\Omega)$ and $B_{ij} : V_h^H(\Omega_{ij}) \to V_h^H(\Omega_{ij})$ be the symmetric and positive definite operators such that

$$(B_k\mathbf{v}, \mathbf{v}) \approxeq (A_k\mathbf{v}_k, \mathbf{v}_k)_{\Omega_k}, \quad \forall \mathbf{v} \in V_h^0(\Omega_k),$$

where $\mathbf{v}_k = \mathbf{v}|_{\Omega_k}$ for $k = 1, 2, \cdots, N$, and

$$(B_d\mathbf{v}_d,\mathbf{v}_d) \eqsim \mathscr{A}(\mathbf{v}_d,\mathbf{v}_d), \quad \forall \mathbf{v}_d \in V_d(\Omega),$$
$$(B_{ij}\mathbf{v},\mathbf{v}) \eqsim \mathscr{A}(\mathbf{v},\mathbf{v}), \quad \forall \mathbf{v} \in V_h^H(\Omega_{ij}).$$

The symbol $\eqsim$ above means each of the two quantities involved is bounded by the other up to a constant independent of h, d and functions involved in the two quantities.

The local solvers on $V_h^{\mathrm{E}}(\Omega)$ should be solvable in an efficient manner, and their constructions are much more tricky and technical than the others. To do so, we introduce more notation.

For any face F from the triangulations $\mathscr{T}_d$, we use F_b to denote the union of all $\mathscr{T}_h$-induced (closed) triangles on F, which have either one single vertex or one edge lying on $\partial\mathrm{F}$, and F_∂ to denote the open set $\mathrm{F}\backslash\mathrm{F}_b$. For any subdomain Ω_k, define

$$\Delta_k = \bigcup_{\mathrm{F}\subset\Gamma_k} \mathrm{F}_b, \quad k = 1,\cdots,N.$$

We will also need the so-called tangential divergence $\mathrm{div}_\tau \Phi = curl_S \Phi$ for $\Phi \in V_h(\Gamma_k)$, which is defined here as in [1, 2]. Then we can introduce our local solver $B_{\mathrm{E}} : V_h^{\mathrm{E}}(\Omega) \to V_h^{\mathrm{E}}(\Omega)$ as follows:

$$(B_{\mathrm{E}}\mathbf{v},\mathbf{u}) = h[1+\log(d/h)] \sum_{k=1}^{N} \Big\{ \alpha_k \langle \mathrm{div}_\tau(\mathbf{v}\times\mathbf{n})|_{\Gamma_k}, \mathrm{div}_\tau(\mathbf{u}\times\mathbf{n})|_{\Gamma_k}\rangle_{\Delta_k} + \beta_k \langle \mathbf{v}\times\mathbf{n}, \mathbf{u}\times\mathbf{n}\rangle_{\Delta_k} \Big\}, \qquad \mathbf{v} \in V_h^{\mathrm{E}}(\Omega),\ \forall \mathbf{u} \in V_h^{\mathrm{E}}(\Omega). \tag{7}$$

For convenience, we call B_{E} an E-basket local solver.

Let $Q_k : V_h(\Omega) \to V_h^0(\Omega_k)$, $Q_d : V_h(\Omega) \to V_d(\Omega)$, $Q_{\mathrm{E}} : V_h(\Omega) \to V_h^{\mathrm{E}}(\Omega)$ and $Q_{ij} : V_h(\Omega) \to V_h^H(\Omega_{ij})$ be the standard L^2-projections. Then we are ready to propose our new preconditioner for A as follows:

$$B^{-1} = B_d^{-1}Q_d + \sum_{k=1}^{N} B_k^{-1}Q_k + \omega \sum_{\mathrm{E}} B_{\mathrm{E}}^{-1}Q_{\mathrm{E}} + \sum_{\Gamma_{ij}} B_{ij}^{-1}Q_{ij}, \tag{8}$$

where ω is a (constant) relaxation parameter, which is introduced to obtain a balance between the local solvers B_{E} and other remaining solvers.

3.2 Algorithm Based on the New Preconditioner and Main Results

The action of the preconditioner B^{-1} which is needed in each PCG iteration can be described in the following algorithm.

Algorithm 4.1. For $\mathbf{g} \in V_h(\Omega)$, we can compute $\mathbf{u} = B^{-1}\mathbf{g}$ in five steps.

Step 1. Solve the system for $\mathbf{u}_d \in V_d(\Omega)$:

$$(B_d\mathbf{u}_d,\mathbf{v}_d) = (\mathbf{g},\mathbf{v}_d), \quad \forall \mathbf{v}_d \in V_d(\Omega);$$

Step 2. Solve the following system for $\mathbf{u}_k \in V_h^0(\Omega_k)$ in each subdomain in parallel:

$$(B_k\mathbf{u}_k, \mathbf{v}) = (\mathbf{g}, \mathbf{v}), \quad \forall \mathbf{v} \in V_h^0(\Omega_k), \ \ k = 1, \cdots, N;$$

Step 3. Solve the following system for $\mathbf{u}_{ij} \in V_h^0(\Omega_{ij})$ in each subdomain Ω_{ij} in parallel:

$$(B_{ij}\mathbf{u}_{ij}, \mathbf{v}) = (\mathbf{g}, \mathbf{v}) - (A_i\mathbf{u}_i, \mathbf{v})_{\Omega_i} - (A_j\mathbf{u}_j, \mathbf{v})_{\Omega_j}, \quad \forall \mathbf{v} \in V_h^0(\Omega_{ij});$$

Step 4. Solve the system for $\mathbf{u}_{\mathrm{E}} \in V_h^{\mathrm{E}}(\Omega)$:

$$(B_{\mathrm{E}}\mathbf{u}_{\mathrm{E}}, \mathbf{v}) = (\mathbf{g}, \tilde{\mathbf{v}}) - \sum_{k=1}^{N} (A_k\mathbf{u}_k, \tilde{\mathbf{v}}), \ \mathbf{v} \in V_h^{\mathrm{E}}(\Omega),$$

where $\tilde{\mathbf{v}} \in V_h(\Omega)$ is a natural extension of $(\mathbf{v} \times \mathbf{n})|_\Gamma$ by zero.
Step 5. Set $\Phi_h = (\sum_{\Gamma_{ij}} \mathbf{u}_{ij} + \sum_{\mathrm{E}} \mathbf{u}_{\mathrm{E}}) \times \mathbf{n}|_\Gamma$ and compute the A-extension of Φ_h on each Ω_k to obtain $\mathbf{u}^H \in V_h^H(\Omega)$. This leads to

$$\mathbf{u} = \mathbf{u}_d + \sum_{k=1}^{N} \mathbf{u}_k + \mathbf{u}^H.$$

Remark 1. For the local solver B_{ij} on each face Γ_{ij}, we may use the face extended domain formed by, e.g., one half of each of the two neighboring subdomains Ω_i and Ω_j. Such definition of B_{ij}'s can reduce the computational complexity in their numerical realization.

Let E denote a coarse edge of the subdomain D_r. Define

$$V_h^{\perp}(\Omega) = \{\mathbf{v}_h : \mathbf{v}_h \in V_h(\Omega), \ \int_E \mathbf{v}_h \cdot \mathbf{t}_E ds = 0 \text{ for each } E\}.$$

We shall use $\kappa^{\perp}(B^{-1}A)$ to denote the *induced condition number* of the preconditioned system $B^{-1}A$ associated with the subspace $V_h^{\perp}(\Omega)$, namely the condition number of $B^{-1}A$ restricted on the subspace $V_h^{\perp}(\Omega)$ (cf. [17]). At this moment we are able to establish only the following estimate of the induced condition number. As the estimate is quite lengthy and technical, we cannot include it here due to the page limitation.

Theorem 1. *Under the assumptions (2), the preconditioner B given in (8) is nearly optimal in the sense that*

$$\kappa^{\perp}(B^{-1}A) \leq C[1 + \log(d/h)]^2[1 + \log(1/h)]^2 \tag{9}$$

where the constant C is independent of h, d and the jumps of the coefficients.

As we see from the above theorem that the induced condition number grows logarithmically with the degrees of freedom in each subdomain, but also with the

degrees of freedom of the entire fine mesh. We believe this is mainly due to the restriction of our current analysis technique, namely the estimate must be done for the induced condition number in the subspace $V_h^{\perp}(\Omega)$ associated with the coarse triangulation formed by the material subdomains D_r. We expect the estimate should be finally carried out directly in the entire edge element space $V_h(\Omega)$, that will remove the logarithmic factor of $1/h$ in the estimate (9). This expectation has already been confirmed by our three-dimensional numerical experiments; see the next section.

4 Numerical Experiments

In this section we shall conduct some numerical experiments to check the convergence of the newly proposed preconditioner, and find out whether they are consistent with the prediction of the convergence theory developed in the previous sections.

In our experiments, we take the domain to be the unit cube $\Omega = (0,1)^3$, while the right-hand side $\mathbf{f}$ of the system (1) is selected such that the exact solution $\mathbf{u} = (u_1, u_2, u_3)^T$ is given by

$$\begin{aligned} u_1 &= xyz(x-1)(y-1)(z-1), \\ u_2 &= \sin(\pi x)\sin(\pi y)\sin(\pi z), \\ u_3 &= (1-e^x)(1-e^{x-1})(1-e^y)(1-e^{y-1})(1-e^z)(1-e^{z-1}), \end{aligned}$$

when the coefficients $\alpha(\mathbf{x})$ and $\beta(\mathbf{x})$ are both constant 1. This right-hand side $\mathbf{f}$ is then fixed in all our experiments, but the coefficients $\alpha(\mathbf{x})$ and $\beta(\mathbf{x})$ may be taken differently.

We then need to triangulate the domain Ω into subdomains $\{\Omega_k\}$. For this, we first partition the three edges of Ω on x-, y- and z-axis into n equal subintervals from which one can naturally generate n^3 equal smaller cubes of size $d = 1/n$. This yields the desired subdomain decomposition in our experiments.

Next, we further triangulate each subdomain Ω_k to get a fine triangulation $\mathscr{T}_h$ of size h over the domain Ω. To generate $\mathscr{T}_h$, we divide each subdomain into m^3 equal smaller cubes of size $h = 1/(mn)$, in the same manner as done in the previous subdomain generation. Then $\mathscr{T}_h$ is obtained by triangulating each cube into six tetrahedra. For easy identification, we may denote the triangulation $\mathscr{T}_h$ as $m^3(n^3)$ below.

The edge finite element space of the lowest order is used for the discretization of (3). The resulting system (5) is solved by PCG method with the newly proposed preconditioners B defined in Sect. 4. We shall choose the balancing parameter ω in front of the E-*basket* local solvers B_{E} in (8) as $\omega = 1$ or $\omega = 2.5$.

We consider various distributions of the coefficients $\alpha(\mathbf{x})$ and $\beta(\mathbf{x})$ and report the corresponding numbers of PCG iterations, and the condition numbers of $B^{-1}A$ for some representative cases. The PCG iteration is terminated in our experiments when the relative residual is less than 10^{-6}.

Case (i): coefficients $\alpha(\mathbf{x}) = \beta(\mathbf{x}) = 1$, with no jumps. The PCG iterations and the condition numbers (in brackets) for $\omega = 2.5$ are listed in Table 1.

	$\omega = 1.0$				$\omega = 2.5$			
$m \backslash n$	4	6	8	10	4	6	8	10
4	34	33	32	32	31 (34.24)	31 (36.31)	31 (36.94)	30 (37.40)
8	41	40	39	38	39 (52.15)	38 (53.78)	37 (54.21)	37 (54.61)
12	48	47	44	42	43 (64.29)	43 (65.91)	41 (66.19)	41 (66.62)
16	51	50	49	45	47 (74.40)	46 (75.69)	44 (75.82)	44 (76.39)

Table 1. Iterations (and condition numbers) with smooth coefficients

We observe from the above table that the number of PCG iterations grows slowly when $m = d/h$ increases but $n = 1/d$ is fixed, and that these numbers vary stably when m is fixed but n increases. This justifies our early expection that the condition number of the preconditioned system $B^{-1}A$ should grow logarithmically with d/h only, not with $1/h$.

One important issue we like to draw the readers' attention to is the large-scale of the discrete system we are solving. For instance, when $m = 16$ and $n = 10$, the total number of degrees of freedom for the fine edge element system is about 28,672,000.

Case (ii): coefficients $\alpha(\mathbf{x})$ and $\beta(\mathbf{x})$ have large jumps:

$$\alpha(\mathbf{x}) = \beta(\mathbf{x}) = \alpha_0 \quad \text{in} \quad D; \quad \alpha(\mathbf{x}) = \beta(\mathbf{x}) = 1 \quad \text{in} \quad \Omega \backslash D.$$

where $D \subset \Omega$ is a union of several subdomains Ω_k. We choose $\alpha_0 = 10^{-5}$ or $\alpha_0 = 10^5$, and consider two choices of D, where one does not have *cross-points*, while the other has one *cross-point*.

Example 1:

$$D = [\frac{1}{4}, \frac{1}{2}]^3.$$

Example 2:

$$D = [\frac{1}{4}, \frac{1}{2}]^3 \bigcup [\frac{1}{2}, \frac{3}{4}]^3.$$

The numerical results are given in Tables 2 and 3, from which we can make some similar observations about the PCG convergence in terms of the mesh and subdomain quantities d/h and d as we did for Case (i).

	Example 1				Example 2			
	$\omega = 1.0$		$\omega = 2.5$		$\omega = 1.0$		$\omega = 2.5$	
$m \backslash n$	4	8	4	8	4	8	4	8
4	29	31	26 (32.00)	29 (35.97)	28	30	26 (35.51)	30 (35.97)
8	35	38	32 (44.88)	37 (52.97)	35	38	32 (45.88)	37 (52.59)
12	38	45	36 (56.02)	42 (64.96)	37	45	35 (55.66)	41 (63.81)
16	40	49	37 (64.65)	45 (74.68)	40	49	37 (65.65)	45 (74.31)

Table 2. Iterations (and condition numbers) with $\alpha_0 = 10^{-5}$

	Example 1				Example 2			
	$\omega = 1.0$		$\omega = 2.5$		$\omega = 1.0$		$\omega = 2.5$	
$m \backslash n$	4	8	4	8	4	8	4	8
4	42	42	36 (40.47)	36 (42.71)	42	44	38 (40.55)	37 (42.72)
8	49	48	45 (61.08)	44 (62.89)	52	51	46 (60.20)	45 (62.89)
12	55	54	50 (74.04)	49 (76.28)	56	56	50 (76.24)	51 (76.28)
16	59	57	54 (91.51)	52 (86.45)	59	59	53 (83.35)	54 (86.45)

Table 3. Iterations (and condition numbers) with $\alpha_0 = 10^5$

Case (iii): coefficients $\alpha(\mathbf{x})$ and $\beta(\mathbf{x})$ have large jumps:

$$\alpha(\mathbf{x}) = \begin{cases} \alpha_0, & in\ D \\ 1, & in\ \Omega \backslash D, \end{cases} \qquad \beta(\mathbf{x}) = \begin{cases} \beta_0, & in\ D \\ 1, & in\ \Omega \backslash D, \end{cases}$$

where $D \subset \Omega$ is a union of several subdomains Ω_k. We choose $\alpha_0 = 10^{-5}$ or $\alpha_0 = 10^5$, but $\beta_0 \neq \alpha_0$. We still consider two different regions D from Examples 1 and 2 in the previous Case (ii), but choose the balancing parameter ω in front of the E-*basket* local solvers B_{E} in (8) as $\omega = 2.5$.

The numerical results are given in Tables 4 and 5. Again, we can make similar observations about the PCG convergence in terms of the mesh and subdomain quantities d/h and d as we did for Case (i).

	Example 1				Example 2			
	$\beta_0 = \alpha_0 \times 10^2$		$\beta_0 = \alpha_0 \times 10^{-2}$		$\beta_0 = \alpha_0 \times 10^2$		$\beta_0 = \alpha_0 \times 10^{-2}$	
$m \backslash n$	4	8	4	8	4	8	4	8
4	30	36	46	47	30	36	45	47
8	39	43	56	56	39	45	56	56
16	49	52	65	65	49	52	63	65

Table 4. Iterations with $\alpha_0 = 10^{-5}$

	Example 1				Example 2			
	$\beta_0 = \alpha_0 \times 10^2$		$\beta_0 = \alpha_0 \times 10^{-2}$		$\beta_0 = \alpha_0 \times 10^2$		$\beta_0 = \alpha_0 \times 10^{-2}$	
$m \backslash n$	4	8	4	8	4	8	4	8
4	31	37	38	41	31	37	39	46
8	37	47	46	49	37	47	53	58
16	48	56	55	57	48	56	66	73

Table 5. Iterations with $\alpha_0 = 10^5$

We may also observe from the previous numerical experiments that appropriate choices of the parameter ω can significantly improve the efficiency of the preconditioner B. It is important to see that the choices of ω seem independent of the fine

and coarse meshsizes h and d, so we may determine ω by solving some small scale systems, e.g., a system with $m = n = 4$.

Acknowledgments QH was supported by the Major Research Plan of Natural Science Foundation of China G91130015, the Key Project of Natural Science Foundation of China G11031006 and National Basic Research Program of China G2011309702. SS was supported by NSFC Project 91130002 and 11171281, the project of Scientific Research Fund of Hunan Provincial Education Department 10C1265 and 11C1219, and the project of Xiangtan University 10XZX03. JZ was substantially supported by Hong Kong RGC grant (Project 405110) and a Direct Grant from Chinese University of Hong Kong.

Bibliography

[1] A. Alonso and A. Valli. Some remarks on the characterization of the space of tangential traces of $H(\mathbf{curl};\Omega)$ and the construction of an extension operator. *Manuscr. Math.*, 89:159–178, 1996.

[2] A. Alonso and A. Valli. An optimal domain decomposition preconditioner for low-frequency time-harmonic maxwell equations. *Math. Comp.*, 68(6):607–631, 1999.

[3] J. Bramble, J. Pasciak, and A. Schatz. The construction of preconditioner for elliptic problems by substructuring. *IV. Math. Comp.*, 53(5):1–24, 1989.

[4] M. Cessenat. *Mathematical methods in electromagnetism*. World Scientific, River Edge, NJ, 1998.

[5] Z. Chen, Q. Du, and J. Zou. Finite element methods with matching and non-matching meshes for maxwell equations with discontinuous coefficients. *SIAM J. Numer. Anal.*, 37:1542–1570, 1999.

[6] P. Ciarlet, Jr., and J. Zou. Fully discrete finite element approaches for time-dependent maxwell's equations. *Numer. Math.*, 82(8):193–219, 1999.

[7] C. R. Dohrmann and O. Widlund. An iterative substructuring algorithm for two-dimensional problems in $H(\mathbf{curl})$. Technical report, TR2010-936, Courant Institute, New York, 2010.

[8] R. Hiptmair. Finite elements in computational electromagnetism. *Acta Numerica*, 11:237–339, 2002.

[9] Q. Hu and J. Zou. A non-overlapping domain decomposition method for maxwell's equations in three dimensions. *SIAM J. Numer. Anal.*, 41:1682–1708, 2003.

[10] Q. Hu and J. Zou. Substructuring preconditioners for saddle-point problems arising from maxwell's equations in three dimensions. *Math. Comput.*, 73:35–61, 2004.

[11] P. Monk. Analysis of a finite element method for maxwell's equations. *SIAM J. Numer. Anal.*, 29:32–56, 1992.

[12] P. Monk. *Finite Element Methods for Maxwell's Equations*. Oxford University Press, Oxford, 2003.

[13] J. Nedelec. Mixed finite elements in R^3. *Numer. Math.*, 35:315–341, 1980.

[14] A. Toselli. Dual-primal FETI algorithms for edge finite-element approximations in 3d. *IMA J. Numer. Anal.*, 26:96–130, 2006.

[15] A. Toselli and O. Widlund. *Domain Decomposition Methods – Algorithms and Theory*. Springer, New York, 2004.

[16] A. Toselli, O. Widlund, and B. Wohlmuth. An iterative substructuring method for Maxwell's equations in two dimensions. *Math. Comp.*, 70:935–949, 2001.

[17] J. Xu and Y. Zhu. Uniform convergent multigrid methods for elliptic problems with strongly discontinuous coefficients. M^3AS, 18:77–105, 2008.

[18] J. Xu and J. Zou. Some non-overlapping domain decomposition methods. *SIAM Review*, 40:857–914, 1998.

Part II

Minisymposia

A Two-Level Schwarz Preconditioner for Heterogeneous Problems

V. Dolean[1], F. Nataf[2], R. Scheichl[3] and N. Spillane[2]

[1] Laboratoire J.A. Dieudonné, CNRS UMR 6621, 06108 Nice Cedex 02, France. dolean@unice.fr
[2] Laboratoire J.L. Lions, CNRS UMR 7598, Université Pierre et Marie Curie, 75005 Paris, France. nataf@ann.jussieu.fr, spillane@ann.jussieu.fr
[3] Department of Mathematical Sciences, University of Bath, Bath BA2 7AY, United Kingdom, R.Scheichl@maths.bath.ac.uk

1 Introduction

Coarse space correction is essential to achieve algorithmic scalability in domain decomposition methods. Our goal here is to build a robust coarse space for Schwarz–type preconditioners for elliptic problems with highly heterogeneous coefficients when the discontinuities are not just across but also along subdomain interfaces, where classical results break down [3, 6, 9, 15].

In previous work, [7], we proposed the construction of a coarse subspace based on the low-frequency modes associated with the Dirichlet-to-Neumann (DtN) map on each subdomain. A rigorous analysis was recently provided in [2]. Similar ideas to build stable coarse spaces, based on the solution of local eigenvalue problems on entire subdomains, can be found in [4], and even traced back to similar ideas for algebraic multigrid methods in [1]. We will argue below that the DtN coarse space presented here is unaffected by coefficient variations that are strictly interior to the subdomain, being as robust as the coarse space analysed in [4].

The robustness result that we obtain, generalizes the classical estimates for overlapping Schwarz methods to the case where the coarse space is richer than just the constant mode per domain [8], or other classical coarse spaces (cf. [15]). The analysis is inspired by that in [4, 13] and crucially uses the framework of weighted Poincaré inequalities, introduced in [10, 11] and successfully applied also to other methods in [12, 14].

2 Two-Level Schwarz Method with DtN Coarse Space

We consider the variational formulation of a second order, elliptic boundary value problem with Dirichlet boundary conditions: Find $u^* \in H_0^1(\Omega)$, for a given domain

R. Bank et al. (eds.), *Domain Decomposition Methods in Science and Engineering XX*, Lecture Notes in Computational Science and Engineering 91,
DOI 10.1007/978-3-642-35275-1_8,

$\Omega \subset \mathbb{R}^d$ ($d = 2$ or 3) and a source term $f \in L_2(\Omega)$, such that

$$a(u^*, v) \equiv \int_\Omega \alpha(x)\, \nabla u^* \cdot \nabla v \;=\; \int_\Omega f v \equiv (f, v)\,, \quad \forall v \in H_0^1(\Omega), \tag{1}$$

and the diffusion coefficient $\alpha = \alpha(x)$ is a positive piecewise constant function that may have large variations within Ω.

We consider a discretization of the variational problem (1) with continuous, piecewise linear finite elements (FE). For a shape regular, simplicial triangulation $\mathscr{T}_h$ of Ω, the standard space of continuous and piecewise linear functions (w.r.t $\mathscr{T}_h$) is then denoted by V_h. The subspace of functions from V_h that vanish on the boundary of Ω is denoted by $V_{h,0}$. The discrete FE problem that we want to solve is: Find $u_h \in V_{h,0}$ such that

$$a(u_h, v_h) = (f, v_h), \quad \forall v_h \in V_{h,0}. \tag{2}$$

Given the usual nodal basis $\{\phi_i\}_{i=1}^n$ for $V_{h,0}$ consisting of "hat" functions with $n := \dim(V_{h,0})$, (2) can be compactly written as

$$A\mathbf{u} \;=\; \mathbf{f}, \quad \text{with} \quad A_{ij} := a(\phi_j, \phi_i) \;\text{ and }\; f_i = (f, \phi_i),\, i, j = 1, \ldots, n, \tag{3}$$

where $\mathbf{u}$ and $\mathbf{f}$ are respectively the vector of coefficients corresponding to the unknown FE function u_h in (2) and to the r.h.s function f.

Two-level Schwarz type methods for (2) are now constructed by choosing an overlapping decomposition $\{\Omega_j\}_{j=1}^J$ of Ω with a subordinate partition of unity $\{\chi_j\}_{j=1}^J$, as well as a suitable coarse subspace $V_H \subset V_{h,0}$. In practice the overlapping subdomains Ω_j can be constructed automatically given the system matrix A by using a graph partitioner, such as METIS, and adding on a number of layers of fine grid elements to the resulting nonoverlapping subdomains. A suitable partition of unity can be constructed from the geometric information of the fine grid. For more details see e.g. [15] or [2]. We assume that each point $x \in \Omega$ is contained in at most N_0 subdomains Ω_j.

The crucial ingredient to obtain robust two-level methods for problems with heterogeneous coefficients is the choice of coarse space $V_H \subset V_{h,0}$. Let us assume for the moment that we have such a space V_H and a restriction operator R_0 from $V_{h,0}$ to V_H and define restriction operators R_j from functions in $V_{h,0}$ to functions in $V_{h,0}(\Omega_j)$, or from vectors in $\mathbb{R}^n$ to vectors in $\mathbb{R}^{\dim V_{h,0}(\Omega_j)}$, by setting $(R_j u)(x_i) = u(x_i)$ for every grid point $x_i \in \Omega_j$. The two-level overlapping additive Schwarz preconditioner for (3) is then simply

$$M_{AS,2}^{-1} = \textstyle\sum_{j=0}^J R_j^T A_j^{-1} R_j \quad \text{where} \quad A_j := R_j A R_j^T,\; j = 0, \ldots, J. \tag{4}$$

In the classical algorithm V_H consists simply of FEs on a coarser triangulation $\mathscr{T}_H$ of Ω and R_H is the canonical restriction from $V_{h,0}$ to V_H, leading to a fully scalable iterative method with respect to mesh/problem size (provided the overlap size is proportional to the coarse mesh size H). However, unfortunately this preconditioner is not robust to strong variations in the coefficient α. We will now present a new,

completely local approach to construct a robust coarse space, as well as an associated restriction operator using eigenvectors of local Dirichlet-to-Neumann maps, proposed in [7].

We start by constructing suitable local functions on each subdomain Ω_j that will then be used to construct a basis for V_H. To this end, let us fix $j \in \{1,\ldots,J\}$ and first consider at the continuous level the Dirichlet-to-Neumann map DtN_j on the boundary of Ω_j. Let $\Gamma_j := \partial\Omega_j$ and let $v_\Gamma : \Gamma_j \to \mathbb{R}$ be a given function, such that $v_\Gamma|_{\partial\Omega} = 0$ if $\Gamma_j \cap \partial\Omega \neq \emptyset$. We define

$$\mathrm{DtN}_j(v_\Gamma) := \alpha \frac{\partial v}{\partial \nu_j}\Big|_{\Gamma_j},$$

where ν_j is the unit outward normal to Ω_j on Γ_j, and v satisfies

$$-\operatorname{div}(\alpha\nabla v) = 0 \text{ in } \Omega_j, \quad v = v_\Gamma \text{ on } \Gamma\,. \tag{5}$$

The function v is the α–harmonic extension of the boundary data v_Γ to the interior of Ω_j.

To construct the (local) coarse basis functions, we now find the low frequency modes of the Dirichlet-to-Neumann operator DtN_j with respect to the weighted L_2–norm on Γ_j, i.e. the smallest eigenvalues of

$$\mathrm{DtN}_j(v_\Gamma^{(j)}) = \lambda^{(j)}\, \alpha v_\Gamma^{(j)}\,. \tag{6}$$

Then we extend each of these modes $v_\Gamma^{(j)}$ α–harmonically to the whole domain and let $v^{(j)}$ be its extension. This is equivalent to the Steklov eigenvalue problem of looking for the pair $(v^{(j)}, \lambda^{(j)})$ which satisfies:

$$-\operatorname{div}(\alpha\nabla v^{(j)}) = 0 \text{ in } \Omega_j \quad \text{and} \quad \alpha\frac{\partial v^{(j)}}{\partial \nu_j} = \lambda\, \alpha v^{(j)} \text{ on } \Gamma_j. \tag{7}$$

The variational formulation of (7) is to find $(v^{(j)}, \lambda^{(j)}) \in H^1(\Omega_j) \times \mathbb{R}$ such that

$$\int_{\Omega_j} \alpha\nabla v^{(j)} \cdot \nabla w = \lambda^{(j)} \int_{\Gamma_j} \mathrm{tr}_j \alpha\, v^{(j)}\, w\,, \quad \forall w \in H^1(\Omega_j), \tag{8}$$

where $\mathrm{tr}_j\alpha(x) := \lim_{y\in\Omega_j \to x} \alpha(y)$. To discretize this generalized eigenvalue problem, we consider for all $v, w \in H^1(\Omega_j)$ the bilinear forms

$$a_j(v,w) := \int_{\Omega_j} \alpha\nabla v \cdot \nabla w \quad \text{and} \quad m_j(v,w) := \int_{\Gamma_j} \mathrm{tr}_j \alpha v w$$

and restrict (8) to the FE space $V_h(\Omega_j)$. The coefficient matrices associated with the variational forms a_j and m_j are

$$A_{kl}^{(j)} := \int_{\Omega_j} \alpha\nabla\phi_k \cdot \nabla\phi_l \quad \text{and} \quad M_{kl}^{(j)} := \int_{\Gamma_j} \mathrm{tr}_j \alpha \phi_k\, \phi_l,$$

where ϕ_k and ϕ_l are any two nodal basis functions for $V_h(\Omega_j)$ associated with vertices of $\mathscr{T}_h$ contained in $\overline{\Omega}_j$. Then the FE approximation to (8) in matrix notation is

$$A^{(j)}\mathbf{v}^{(j)} = \lambda^{(j)} M^{(j)}\mathbf{v}^{(j)} \tag{9}$$

where $\mathbf{v}^{(j)} \in \mathbb{R}^{n_j}$, $n_j := \dim V_h(\Omega_j)$, denotes the degrees of freedom of the FE approximation to $v^{(j)}$ in $V_h(\Omega_j)$.

Let the n_j eigenpairs $(\lambda_\ell^{(j)}, \mathbf{v}_\ell)_{\ell=1}^{n_j}$ corresponding to (9) be numbered in increasing order of $\lambda_\ell^{(j)}$. Since $M_{kl}^{(j)} \neq 0$ only if ϕ_k and ϕ_l are associated with the n_Γ vertices of $\mathscr{T}_h$ that lie on Γ_j, it is easy to see that at most n_Γ of the eigenvalues $\lambda_\ell^{(j)}$ are finite. Moreover, the smallest eigenvalue $\lambda_1^{(j)} = 0$ with constant eigenvector and the set of eigenvectors $\{\mathbf{v}_\ell\}_{\ell=1}^{n_j}$ can be chosen so that they are $A^{(j)}$–orthonormal. The local coarse space is now defined as the span of the FE functions $v_\ell^{(j)} \in V_h(\Omega_j)$, $\ell \leq m_j \leq n_\Gamma$, corresponding to the first m_j eigenpairs of (9). For each subdomain Ω_j, we choose the value of m_j such that $\lambda_\ell^{(j)} < \mathrm{diam}(\Omega_j)^{-1}$, for all $\ell \leq m_j$, and $\lambda_{m_j+1}^{(j)} \geq \mathrm{diam}(\Omega_j)^{-1}$. We will see in the analysis in the next section why this is a sensible choice.

Using the partition of unity $\{\chi_j\}_{j=1}^J$, we now combine the local basis functions constructed in the previous section to obtain a conforming coarse space $V_H \subset V_{h,0}$ on all of Ω. The new coarse space is defined as

$$V_H := \mathrm{span}\left\{ I_h\left(\chi_j v_\ell^{(j)}\right) : 1 \leq j \leq J \text{ and } 1 \leq \ell \leq m_j \right\}, \tag{10}$$

where I_h is the standard nodal interpolant onto $V_{h,0}(\Omega)$. The dimension of V_H is $\sum_{j=1}^J m_j$. By construction each of the functions $I_h\left(\chi_j v_\ell^{(j)}\right) \in V_{h_0}$, so that as required $V_H \subset V_{h,0}$. The transfer operator R_0 from V_{h_0} to V_H is defined in a canonical way by setting $R_0^T u_H(x_i) = u_H(x_i)$, for all $u_H \in V_H$ and for all vertices x_i of $\mathscr{T}_h$.

We will see in the next section that under some mild assumptions on the variability of α this choice of coarse space leads to a scalable and coefficient-robust domain decomposition method with supporting theory.

3 Conditioning Analysis

To analyse this method let us first define the boundary layer $\Omega_j^\circ := \{x \in \Omega_j : \chi_j(x) < 1\}$ for each Ω_j that is overlapped by neighbouring domains, i.e. We assume that this layer is uniformly of width $\geq \delta_j$, in the sense that it can be subdivided into shape regular regions of diameter δ_j, and that the triangulation $\mathscr{T}_h$ resolves it. This also guarantees that it is possible to find a partition of unity such that $|\chi_j| = \mathscr{O}(1)$ and $|\nabla\chi_j| = \mathscr{O}(\delta_j^{-1})$.

We now state the key assumption on the coefficient distribution $\alpha(x)$.

Assumption 1 We assume that, for each $j = 1, \ldots, J$, there exists a set $X_j \subset \Gamma_j$ (not necessarily connected) such that (i) $\max_{x,y\in X_k} \frac{\alpha(x)}{\alpha(y)} = \mathscr{O}(1)$ and (ii) there exists a path

P_y from each $y \in \Omega_j$ to X_j, such that $\alpha(x)$ is an increasing function along P_y (from y to X_j).

Lemma 1 (weighted Poincaré inequality [10]). *Let Assumption 1 hold.*

$$\int_{\Omega_j^\circ} \alpha |v - \bar{v}^{X_j}|^2 \;\le\; C_P\, \delta_j \int_{\Omega_j^\circ} \alpha |\nabla v|^2, \qquad \text{for all } \; v \in V_h(\Omega_j),$$

where $\bar{v}^{X_j} := \frac{1}{|X_j|} \int_{X_j} v$.

Remark 1. Note that Assumption 1 is related to the classical notion of quasi-monotonicity coined in [3]. It ensures that the constant C_P in the Poincaré-type inequality in Lemma 1, as well as all the other (hidden) constants below are independent of the values of the coefficient function $\alpha(x)$. The constants may however depend logarithmically or linearly on δ_j/h. This depends on the geometry and shape of the paths P_y and on the size and shape of the set X_j. For more details see [2] and [10, 11].

The following proposition [2, Theorem 3.1] is the central result in our analysis. It proves the stability and a weak approximation property for a local projection onto the span of the first m_j eigenvectors.

Proposition 1. *Let Assumption 1 hold, and for any $u \in V_h(\Omega_j)$, define the projection $\Pi_j u := \sum_{\ell=1}^{m_j} a_j(v_\ell^{(j)}, u)\, v_\ell^{(j)}$. Then*

$$|\Pi_j u|_{a,\Omega_j} \;\le\; |u|_{a,\Omega_j} \quad \text{and} \tag{11}$$

$$\|u - \Pi_j u\|_{0,\alpha,\Omega_j^\circ} \;\lesssim\; \sqrt{c_j(m_j)}\, \delta_j\, |u|_{a,\Omega_j}\,. \tag{12}$$

where $c_j(m_j) := C_P^2 + \left(\delta_j \lambda_{m_j+1}^{(j)}\right)^{-1}$.

As usual (cf. [15]), the following condition number bound can then be obtained via abstract Schwarz theory by constructing a stable splitting.

Theorem 1. *Let Assumption 1 be satisfied. Then the condition number of the two-level Schwarz algorithm with the coarse space V_H based on local DtN maps and defined in (10) can be bounded by*

$$\kappa(M_{AS,2}^{-1} A) \;\lesssim\; \max_{j=1}^{J} \{c_j(m_j)\} \;\lesssim\; C_P^2 + \max_{j=1}^{J} \left(\delta_j \lambda_{m_j+1}^{(j)}\right)^{-1}.$$

The hidden constant is independent of h, δ_j, $\operatorname{diam}(\Omega_j)$, and α.

Proof. We construct a stable splitting for a function $u \in V_{h,0}$ using the projections Π_j, $j = 1, \ldots, J$, in Proposition 1 to define the coarse quasi-interpolant

$$u_0 := I_h\Big(\sum_{j=1}^{J} \chi_j \Pi_j u|_{\Omega_j}\Big) \in V_H\,. \tag{13}$$

If we now choose $u_j := I_h(\chi_j(u - \Pi_j u)) \in V_{h,0}(\Omega_j)$, then

$$u = \sum_{j=0}^{J} u_j \quad \text{and} \quad \sum_{j=0}^{J} \int_\Omega \alpha |\nabla u_j|^2 \;\lesssim\; \max_{j=1}^{J} \{c_j(m_j)\} \int_\Omega \alpha |\nabla u|^2$$

For details see the proof of [2, Theorem 3.3].

Remark 2. Note that by choosing the number m_j of modes per subdomain such that $\lambda_{m_j+1}^{(j)} \geq \mathrm{diam}(\Omega_j)^{-1}$, as stated in Sect. 2, we have

$$\kappa(M_{AS,1}^{-1}A) \lesssim \left(C_P^2 + \max_j \mathrm{diam}(\Omega_j)/\delta_j\right).$$

Hence, provided the constant C_P is uniformly bounded, independently of any jumps in the coefficients, we retrieve the classical estimate for the two-level additive Schwarz method independently of any variations of coefficients across or along subdomain boundaries.

4 Numerical Results

We choose $\Omega = (0,1)^2$ and discretize (1) on a uniform grid with $2m^2$ elements, setting $u = 0$ on the left hand boundary and $\frac{\partial u}{\partial \nu} = 0$ on the remainder. We use METIS to split the domain into 16 irregular subdomains as shown in Fig. 1 and construct the overlapping partition by extending each subdomain by one layer of fine grid elements using Freefem++ [5].

As the coarse space we use the DtN coarse space described in Sect. 2 with m_j chosen such that $\lambda_{m_j}^{(j)} < \mathrm{diam}(\Omega_j)^{-1} \leq \lambda_{m_j+1}^{(j)}$, for all $j = 1, \ldots, 16$ (labelled `D2N`). We compare this preconditioner with the one-level additive Schwarz method (labelled `NONE`) and the two-level method with partition of unity coarse space, i.e. choosing $m_j = 1$ for all j (labelled `POU`). To confirm in some sense the optimality of our choice for m_j, we also include results with the DtN coarse space choosing $m_j + 1$ and $\max\{1, m_j - 1\}$ basis functions per subdomain (labelled `D2N+` and `D2N-`, respectively). We use the preconditioners within a conjugate gradient iteration with tolerance 10^{-7}.

In the first test case (**Example 1**), we choose $m = 160$ and α as depicted in Fig. 2, i.e. 25 high permeability inclusions and one channel. In the second test case (**Example 2**), we choose $m = 80$ and α to be a realization of a log-normal distribution with exponential covariance function (variance $\sigma^2 = 4$ and correlation length $\lambda = 4/m$) and mean of $\log \alpha$ equal 3 (cf. Fig. 3).

In Fig. 4 we plot $\|u - \bar{u}\|_\infty$ for Example 1 against the iteration count, where $\bar{u}$ is the solution of (3) obtained via a direct solver. Clearly both the one-level and the two-level preconditioner with POU coarse space are not robust. The POU coarse space seems to have hardly any influence at all (520 versus 619 iterations), whereas the new DtN coarse space leads to a robust convergence and a significantly reduced number of iterations of 64.

Finally, in Table 1 we compare the different preconditioners and show that the criterion for the number m_j of eigenmodes that we select in each subdomain is in some sense optimal. Adding one more functions has hardly any impact on the performance while removing one has a strong negative impact. See [2] for more extensive numerical experiments.

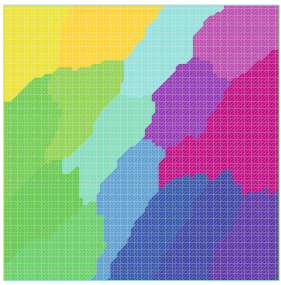


Fig. 1. Partition into 16 subdomains

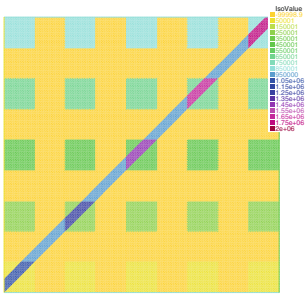


Fig. 2. Example 1 ($\max_{x,y} \frac{\alpha(x)}{\alpha(y)} = 2 \cdot 10^6$)

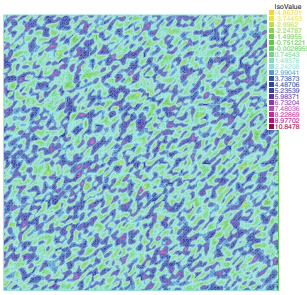


Fig. 3. Example 2 ($\max_{x,y} \frac{\alpha(x)}{\alpha(y)} = 7 \cdot 10^6$)

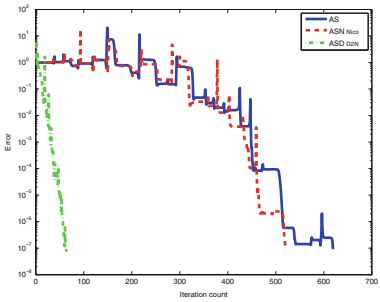


Fig. 4. Convergence history (Example 1)

	Coarse space size dimV_H					# PCG Iterations (tol $= 10^{-7}$)				
	NONE	POU	D2N-	D2N	D2N+	NONE	POU	D2N-	D2N	D2N+
Example 1	0	16	32	46	62	619	520	446	64	37
Example 2	0	16	82	98	114	89	92	50	38	36

Table 1. Comparison of DtN coarse space against simple POU coarse space and no coarse space, as well as demonstration of "optimality" of automatic criterion for choosing $\{m_j\}$.

Bibliography

[1] T. Chartier, R. D. Falgout, V. E. Henson, J. Jones, T. Manteuffel, S. McCormick, J. Ruge, and P. S. Vassilevski. Spectral AMGe (ρAMGe). *SIAM J. Sci. Comput.*, 25(1):1–26, 2003. ISSN 1064-8275. doi: 10.1137/S106482750139892X. URL http://dx.doi.org/10.1137/S106482750139892X.

[2] V. Dolean, F. Nataf, Scheichl R., and N. Spillane. Analysis of a two-level Schwarz method with coarse spaces based on local Dirichlet–to–Neumann maps. http://hal.archives-ouvertes.fr/hal-00586246/fr/, 2011. URL http://hal.archives-ouvertes.fr/hal-00586246/fr/.

[3] M. Dryja, M. V. Sarkis, and O. B. Widlund. Multilevel Schwarz methods for elliptic problems with discontinuous coefficients in three dimensions. *Numer.*

Math., 72(3):313–348, 1996. ISSN 0029-599X. doi: 10.1007/s002110050172. URL http://dx.doi.org/10.1007/s002110050172.

[4] J. Galvis and Y. Efendiev. Domain decomposition preconditioners for multiscale flows in high contrast media: Reduced dimension coarse spaces. *Multiscale Modeling & Simulation*, 8(5):1621–1644, 2010. doi: 10.1137/100790112.

[5] Frédéric Hecht. *FreeFem++*. Laboratoire J.L. Lions, Université Pierre et Marie Curie, http://www.freefem.org/ff++/, 3.7 edition, 2010.

[6] J. Mandel and M. Brezina. Balancing domain decomposition for problems with large jumps in coefficients. *Math. Comp.*, 65:1387–1401, 1996. ISSN 0025-5718. doi: 10.1090/S0025-5718-96-00757-0. URL http://dx.doi.org/10.1090/S0025-5718-96-00757-0.

[7] F. Nataf, H. Xiang, and V. Dolean. A two level domain decomposition preconditioner based on local Dirichlet-to-Neumann maps. *C. R. Mathématique*, 348 (21–22):1163–1167, 2010.

[8] R. A. Nicolaides. Deflation of conjugate gradients with applications to boundary value problems. *SIAM J. Numer. Anal.*, 24(2):355–365, 1987. ISSN 0036-1429. doi: 10.1137/0724027. URL http://dx.doi.org/10.1137/0724027.

[9] C. Pechstein and R. Scheichl. Scaling up through domain decomposition. *Appl. Anal.*, 88(10–11):1589–1608, 2009. ISSN 0003-6811. doi: 10.1080/00036810903157204. URL http://dx.doi.org/10.1080/00036810903157204.

[10] C. Pechstein and R. Scheichl. Weighted Poincaré inequalities. Technical Report NuMa-Report 2010-10, Institute of Computational Mathematics, Johannes Kepler University, Linz, December 2010. submitted.

[11] C. Pechstein and R. Scheichl. Weighted Poincaré inequalities and applications in domain decomposition. In Y. Huang, R. Kornhuber, O. Widlund, and J. Xu, editors, *Domain Decomposition Methods in Science and Engineering XIX*, volume 78 of *LNCSE*, pages 197–204. Springer, 2011.

[12] C. Pechstein and R. Scheichl. Analysis of FETI methods for multiscale PDEs - Part II: Interface variation. *Numer. Math.*, 2011. Published online 21 February 2011.

[13] R. Scheichl, P. S. Vassilevski, and L. T. Zikatanov. Weak approximation properties of elliptic projections with functional constraints. Technical Report LLNL-JRNL-462079, Lawrence Livermore National Lab, 2011.

[14] R. Scheichl, P.S. Vassilevski, and L.T. Zikatanov. Multilevel methods for elliptic problems with highly varying coefficients on non-aligned coarse grids. *SIAM J Numer Anal*, 2011. Accepted subject to minor corrections.

[15] A. Toselli and O. B. Widlund. *Domain decomposition methods – algorithms and theory*. Springer, Berlin, 2005. ISBN 3-540-20696-5.

Heterogeneous Domain Decomposition Methods for Eddy Current Problems

Ana Alonso Rodríguez

Dipartimento di Matematica, Università di Trento, I-38050 Povo (Trento), Italy
alonso@science.unitn.it

Summary. The usual setting of an eddy current problem distinguishes between a conducting region and an air region (non-conducting) surrounding the conductor. For the numerical approximation of this heterogeneous problem it is very natural to use iterative substructuring methods based on transmission conditions at the interface. We analyze the convergence of the Dirichlet-Neumann iterative method for two different formulations of the eddy current problem: the one that consider as main unknown the electric field and the one based on the magnetic field.

1 Introduction

To model the electromagnetic phenomena concerning alternating currents at low frequencies it is often used the time-harmonic eddy current model (see e.g. [2]). The main equations of this model are Faraday's law

$$\operatorname{curl} \mathbf{E} = -i\omega\mu\mathbf{H} \quad \text{in } \Omega\,, \tag{1}$$

and Ampère's law

$$\operatorname{curl} \mathbf{H} = \sigma\mathbf{E} + \mathbf{J}_e \quad \text{in } \Omega\,, \tag{2}$$

where $\mathbf{E}$, $\mathbf{H}$ and $\mathbf{J}_e$ denote the electric field, the magnetic field and the applied current density respectively. For the sake of simplicity we assume that the computational domain $\Omega \subset \mathbb{R}^3$ is a simply connected Lipschitz polyhedron with connected boundary that contains a conducting region $\Omega_C \subset\subset \Omega$ and that both Ω_C and its complement $\Omega_I := \Omega \setminus \overline{\Omega_C}$ are connected Lipschitz polyhedra. Let us denote $\Gamma := \overline{\Omega_C} \cap \overline{\Omega_I}$. The magnetic permeability μ is assumed to be a symmetric uniformly positive definite 3×3 matrix with entries in $L^\infty(\Omega)$, whereas the electric conductivity σ is supposed to be a bounded symmetric positive definite matrix in the conducting regions, and to be null in non-conducting regions. The real scalar constant $\omega \neq 0$ is a given angular frequency. In $\partial\Omega$ suitable boundary conditions must be assigned. Most often the tangential component of either the electric field $\mathbf{E} \times \mathbf{n}$ or the magnetic field $\mathbf{H} \times \mathbf{n}$ are given (here $\mathbf{n}$ denotes the unit outward normal vector on $\partial\Omega$).

R. Bank et al. (eds.), *Domain Decomposition Methods in Science and Engineering XX*, Lecture Notes in Computational Science and Engineering 91,
DOI 10.1007/978-3-642-35275-1_9,

Let us introduce some notations that will be used in the following. The space $H(\mathrm{curl}\,;\Omega)$ indicates the set of real or complex vector valued functions $\mathbf{v} \in (L^2(\Omega))^3$ such that curl $\mathbf{v} \in (L^2(\Omega))^3$ and $H^0(\mathrm{curl}\,;\Omega)$ its subspace constituted by curl-free functions. Given a certain subset $\Lambda \subset \partial\Omega$, we denote by $H_{0,\Lambda}(\mathrm{curl}\,;\Omega)$ the subspace of functions in $H(\mathrm{curl}\,;\Omega)$ such that their tangential trace is null on Λ, and in particular we write $H_0(\mathrm{curl}\,;\Omega) := H_{0,\partial\Omega}(\mathrm{curl}\,;\Omega)$.

We recall the spaces $H^{-1/2}(\mathrm{curl}_\tau;\partial\Omega) := \left\{(\mathbf{n}\times\mathbf{v}\times\mathbf{n})_{|\partial\Omega}\,|\,\mathbf{v}\in H(\mathrm{curl}\,;\Omega)\right\}$, and $H^{-1/2}(\mathrm{div}_\tau;\partial\Omega) := \left\{(\mathbf{v}\times\mathbf{n})_{|\partial\Omega}\,|\,\mathbf{v}\in H(\mathrm{curl}\,;\Omega)\right\}$, (see [4]). These two spaces are in duality and the following formula of integration by parts holds true

$$\int_\Omega (\mathbf{w}\cdot\mathrm{curl}\,\overline{\mathbf{v}} - \mathrm{curl}\,\mathbf{w}\cdot\overline{\mathbf{v}}) = \langle \mathbf{w}\times\mathbf{n}, \mathbf{n}\times\overline{\mathbf{v}}\times\mathbf{n}\rangle_{\partial\Omega} \quad \forall\,\mathbf{w},\,\mathbf{v}\in H(\mathrm{curl}\,;\Omega)\,.$$

2 One Field Formulations

First we notice that Eqs. (1) and (2) do not completely determine the electric field in Ω_I and it is necessary to require the gauge condition

$$\mathrm{div}\mathbf{E}_I = 0 \ \text{ in } \Omega_I\,. \tag{3}$$

(Here and in the sequel, given any vector field $\mathbf{v}$ defined in Ω, we denote $\mathbf{v}_L$ its restriction to Ω_L, $L = C, I$.) When imposing electric boundary conditions, $\mathbf{E}\times\mathbf{n} = \mathbf{0}$ on $\partial\Omega$, in order to have a unique solution we need to impose the additional gauge condition $\int_\Gamma \mathbf{E}_I\cdot\mathbf{n} = 0$.

From Faraday law $\mu^{-1}\mathrm{curl}\,\mathbf{E} = -i\omega\mathbf{H}$ and replacing in Ampère law one has curl $(\mu^{-1}\mathrm{curl}\,\mathbf{E}) = -i\omega(\sigma\mathbf{E}+\mathbf{J}_e)$. So the $\mathbf{E}$-based formulation of the eddy current problem with electric boundary conditions reads

$$\begin{array}{ll}
\mathrm{curl}\,(\mu^{-1}\mathrm{curl}\,\mathbf{E}) + i\omega\sigma\mathbf{E} = -i\omega\mathbf{J}_e & \text{in } \Omega \\
\mathrm{div}\mathbf{E}_I = 0 & \text{in } \Omega_I \\
\int_\Gamma \mathbf{E}_I\cdot\mathbf{n} = 0 & \\
\mathbf{E}\times\mathbf{n} = \mathbf{0} & \text{on } \partial\Omega\,.
\end{array}$$

Since $\sigma \equiv 0$ in the non-conducting region, the generator current has to satisfy the compatibility conditions $\mathrm{div}\mathbf{J}_{e,I} = 0$ in Ω_I and, when imposing $\mathbf{E}\times\mathbf{n} = 0$ on $\partial\Omega$, $\int_\Gamma \mathbf{J}_{e,I}\cdot\mathbf{n} = 0$.

Notice that the two gauge conditions $\mathrm{div}\mathbf{E}_I = 0$ and $\int_\Gamma \mathbf{E}_I\cdot\mathbf{n} = 0$ are equivalent to $\int_{\Omega_I} \mathbf{E}_I\cdot\nabla\overline{\phi}_I = 0$ for all $\phi_I \in H^1_*(\Omega_I)$ being $H^1_*(\Omega_I) = \{\phi_I \in H^1(\Omega_I) : \phi_{I|\partial\Omega} \equiv 0$ and $\phi_{I|\Gamma}$ is constant$\}$. Hence the weak form of the $\mathbf{E}$-based formulation is

$$\begin{array}{l}
\text{Find } \mathbf{E}\in W \text{ such that} \\
\quad \int_\Omega(\mu^{-1}\mathrm{curl}\,\mathbf{E}\cdot\mathrm{curl}\,\overline{\mathbf{w}} + i\omega\sigma\mathbf{E}\cdot\overline{\mathbf{w}}) = -i\omega\int_\Omega \mathbf{J}_e\cdot\overline{\mathbf{w}} \\
\text{for all } \mathbf{w}\in W
\end{array}$$

where $W := \{\mathbf{w}\in H_0(\mathrm{curl}\,;\Omega) : \int_{\Omega_I}\mathbf{w}_I\cdot\nabla\overline{\phi}_I = 0\ \forall\,\phi_I\in H^1_*(\Omega_I)\}$.

Remark 1. The gauge conditions can be imposed by means of a Lagrange multiplier. (See [2], Sect. 4.6.)

Due to the heterogeneous nature of the problem, it is natural to consider an iterative procedure by subdomains in order to deal with homogeneous problem. A procedure of this kind is the following:

Given $\boldsymbol{\lambda}^{(0)} \in H^{-1/2}(\text{curl}_{\,\tau};\Gamma)$ for $n \geq 0$

find $\mathbf{E}_I^{(n+1)} \in W_I$ such that
$\mathbf{n} \times \mathbf{E}_I^{(n+1)} \times \mathbf{n} = \boldsymbol{\lambda}^{(n)}$ on Γ
$\int_{\Omega_I} \mu^{-1} \text{curl}\, \mathbf{E}_I^{(n+1)} \cdot \text{curl}\, \overline{\mathbf{w}}_I = -i\omega \int_{\Omega_I} \mathbf{J}_{e,I} \cdot \overline{\mathbf{w}}_I \quad \forall \mathbf{w}_I \in W_I \cap H_0(\text{curl}\,;\Omega_I)$;

find $\mathbf{E}_C^{(n+1)} \in H(\text{curl}\,;\Omega_C)$ such that
$\int_{\Omega_C} (\mu^{-1} \text{curl}\, \mathbf{E}_C^{(n+1)} \cdot \text{curl}\, \overline{\mathbf{w}}_C + i\omega\sigma \mathbf{E}_C^{(n+1)} \cdot \overline{\mathbf{w}}_C) = -i\omega \int_{\Omega_C} \mathbf{J}_{e,C} \cdot \overline{\mathbf{w}}_C$
$\quad - \langle \mu^{-1} \text{curl}\, \mathbf{E}_I^{(n+1)} \times \mathbf{n}_I, \mathbf{n} \times \mathbf{w}_C \times \mathbf{n} \rangle_\Gamma \quad \forall \mathbf{w}_C \in H(\text{curl}\,;\Omega_C)$;

set
$\boldsymbol{\lambda}^{(n+1)} = (1-\theta)\boldsymbol{\lambda}^{(n)} + \theta(\mathbf{n} \times \mathbf{E}_C^{(n+1)} \times \mathbf{n})_{|\Gamma}$,

where $W_I := \{\mathbf{w}_I \in H_{0,\partial\Omega}(\text{curl}\,;\Omega_I) : \int_{\Omega_I} \mathbf{w}_I \cdot \nabla\overline{\phi}_I = 0 \; \forall \phi_I \in H^1_*(\Omega_I)\}$, $\mathbf{n}_I$ denotes the unit normal vector on Γ pointing outwards Ω_I and θ is a positive acceleration parameter.

Another possibility is to eliminate the electric field. Multiplying Faraday law by a function $\mathbf{v} \in H_0(\text{curl}\,;\Omega)$ with curl $\mathbf{v}_I = 0$;

$$\begin{aligned} i\omega \int_\Omega \mu \mathbf{H} \cdot \overline{\mathbf{v}} &= -\int_\Omega \text{curl}\, \mathbf{E} \cdot \overline{\mathbf{v}} = -\int_\Omega \mathbf{E} \cdot \text{curl}\, \overline{\mathbf{v}} \\ &= -\int_{\Omega_C} \sigma^{-1} (\text{curl}\, \mathbf{H}_C - \mathbf{J}_{e,C}) \cdot \text{curl}\, \overline{\mathbf{v}}_C \,. \end{aligned}$$

Given $\mathbf{g}_I \in (L^2(\Omega_I))^3$ let $V(\mathbf{g}_I)$ denotes the space $V(\mathbf{g}_I) := \{\mathbf{v} \in H_0(\text{curl}\,;\Omega) : \text{curl}\, \mathbf{v}_I = \mathbf{g}_I\}$. The weak form of $\mathbf{H}$-based formulation of the eddy current problem with magnetic boundary conditions $\mathbf{H} \times \mathbf{n} = \mathbf{0}$ on $\partial\Omega$ reads

$$\begin{aligned} &\text{Find } \mathbf{H} \in V(\mathbf{J}_{e,I}) \text{ such that} \\ &\quad \int_{\Omega_C} \sigma^{-1} \text{curl}\, \mathbf{H} \cdot \text{curl}\, \overline{\mathbf{v}} + i\omega \int_\Omega \mu \mathbf{H} \cdot \overline{\mathbf{v}} = \int_{\Omega_C} \sigma^{-1} \mathbf{J}_{e,C} \cdot \text{curl}\, \overline{\mathbf{v}}_C \\ &\text{for all } \mathbf{v} \in V(\mathbf{0}) \,. \end{aligned} \tag{4}$$

Since $\sigma \equiv 0$ in the non-conducting region, when imposing $\mathbf{H} \times \mathbf{n} = 0$ on $\partial\Omega$ the generator current has to satisfy the compatibility conditions $\text{div}\mathbf{J}_{e,I} = 0$ in Ω_I and $\mathbf{J}_{e,I} \cdot \mathbf{n} = 0$ on $\partial\Omega$. Hence there exists $\mathbf{H}^*_{e,I} \in H_{0,\partial\Omega}(\text{curl}\,;\Omega_I)$ such that curl $\mathbf{H}^*_{e,I} = \mathbf{J}_{e,I}$. Then we can write $\mathbf{H}_I = \mathbf{H}^*_{e,I} + \mathbf{Z}_I$ with $\mathbf{Z}_I \in H^0_{0,\partial\Omega}(\text{curl}\,;\Omega_I)$. Let $\mathbf{H}^*_e$ be a function in $H(\text{curl}\,;\Omega)$ such that $\mathbf{H}^*_{e|\Omega_I} = \mathbf{H}^*_{e,I}$ and let us denote $\mathbf{Z} := \mathbf{H} - \mathbf{H}^*_e \in V(\mathbf{0})$. Multiplying Eq. (4) by $-i\omega^{-1}$ and setting $\widehat{F}(\mathbf{v}) := \int_\Omega \mu \mathbf{H}^*_e \cdot \overline{\mathbf{v}} - i\omega^{-1} \int_{\Omega_C} \sigma^{-1} \text{curl}\, \mathbf{H}^*_e \cdot \text{curl}\, \overline{\mathbf{v}}$, we can consider the equivalent problem

Find $\mathbf{Z} \in V(\mathbf{0})$ such that
$\int_\Omega \mu \mathbf{Z} \cdot \overline{\mathbf{v}} - i\omega^{-1} \int_{\Omega_C} \sigma^{-1} \mathrm{curl}\, \mathbf{Z} \cdot \mathrm{curl}\, \overline{\mathbf{v}} = -i\omega^{-1} \int_{\Omega_C} \sigma^{-1} \mathbf{J}_{e,C} \cdot \mathrm{curl}\, \overline{\mathbf{v}}_C - \widehat{F}(\mathbf{v})$
for all $\mathbf{v} \in V(\mathbf{0})$.

For the sake of simplicity we will assume that $\mathbf{J}_{e,I} \cdot \mathbf{n} = 0$ on Γ. Then it is possible to take $\mathbf{H}^*_{e,I} \in H_0(\mathrm{curl}\,;\Omega_I)$ and $\mathbf{H}^*_{e,C}$ equal zero.

Remark 2. Notice that $H^0_{0,\partial\Omega}(\mathrm{curl}\,;\Omega_I) = \nabla H^1_{0,\partial\Omega}(\Omega_I) \oplus \mathscr{H}(\Omega_I)$ where $\mathscr{H}(\Omega_I) := \{\mathbf{v}_I \in H^0_{0,\partial\Omega}(\mathrm{curl}\,;\Omega_I) : \mathrm{div}\mathbf{v}_I = 0 \text{ and } \mathbf{v}_I \cdot \mathbf{n} = 0 \text{ on } \Gamma\}$ that is a space of finite dimension. In this geometrical setting the dimension of $\mathscr{H}(\Omega_I)$ coincides with the first Betti number of Ω_I. (See [2], Sect. 5.1.)

We propose an iterative procedure for the solution of the **H**-based formulation that start from a data in the trace space

$$H_0^{-1/2}(\mathrm{curl}\,_\tau;\Gamma) := \{(\mathbf{n} \times \mathbf{w}_I \times \mathbf{n})_{|\Gamma} : \mathbf{w}_I \in H^0_{0,\partial\Omega}(\mathrm{curl}\,;\Omega_I)\}.$$

It reads:

Given $\boldsymbol{\lambda}^{(0)} \in H_0^{-1/2}(\mathrm{curl}\,_\tau;\Gamma)$ for $n \geq 0$

find $\mathbf{H}_C^{(n+1)} \in H(\mathrm{curl}\,;\Omega_C)$ such that
$\mathbf{n} \times \mathbf{H}_C^{(n+1)} \times \mathbf{n} = \boldsymbol{\lambda}^{(n)}$ on Γ
$\int_{\Omega_C} (\mu \mathbf{H}_C^{(n+1)} \cdot \overline{\mathbf{v}}_C - i\omega^{-1}\sigma^{-1} \mathrm{curl}\, \mathbf{H}_C^{(n+1)} \cdot \mathrm{curl}\, \overline{\mathbf{v}}_C)$
$= -i\omega^{-1} \int_{\Omega_C} \sigma^{-1} \mathbf{J}_{e,C} \cdot \mathrm{curl}\, \overline{\mathbf{v}}_C \quad \forall \mathbf{v}_C \in H_0(\mathrm{curl}\,;\Omega_C)$;

find $\mathbf{Z}_I^{(n+1)} \in H^0_{0,\partial\Omega}(\mathrm{curl}\,;\Omega_I)$ such that
$\int_{\Omega_I} \mu \mathbf{Z}_I^{(n+1)} \cdot \overline{\mathbf{v}}_I = i\omega^{-1} \langle \sigma^{-1}(\mathrm{curl}\, \mathbf{H}_C^{(n+1)} - \mathbf{J}_{e,C}) \times \mathbf{n}_C, \mathbf{n} \times \mathbf{v}_I \times \mathbf{n} \rangle_\Gamma$
$- \int_{\Omega_I} \mu \mathbf{H}^*_{e,I} \cdot \overline{\mathbf{v}}_I \quad \forall \mathbf{v}_{I,h} \in H^0_{0,\partial\Omega}(\mathrm{curl}\,;\Omega_I)$;

set
$\boldsymbol{\lambda}^{(n+1)} = (1-\theta)\boldsymbol{\lambda}^{(n)} + \theta(\mathbf{n} \times \mathbf{Z}_I^{(n+1)} \times \mathbf{n})_{|\Gamma}$,

being $\mathbf{n}_C$ the unit normal vector on Γ pointing outwards Ω_C and θ a positive acceleration parameter.

3 Convergence Analysis

Both the **H**-based formulation and the **E**-based formulation are of the form: find $\mathbf{u} \in V \subset H(\mathrm{curl}\,;\Omega)$ such that

$$a(\mathbf{u},\mathbf{v}) = F(\mathbf{v}) \quad \forall \mathbf{v} \in V, \tag{5}$$

where $a(\cdot,\cdot)$ is a sesquilinear form continuous and coercive in $V \times V$ and $F(\cdot)$ is a continuous linear functional on the Hilbert space V. The proposed iterative

procedures are preconditioned Richardson methods for the Steklov-Poincare equation obtained in the following way (see e.g. [8]): for $L = C, I$ let us define the spaces $V_L := \{\mathbf{v}_{|\Omega_L} : \mathbf{v} \in V\}$, $X := \{(\mathbf{n} \times \mathbf{v} \times \mathbf{n})_\Gamma : \mathbf{v} \in V\}$ and $V_{L,0} := \{\mathbf{v}_L \in V_L : (\mathbf{n} \times \mathbf{v}_L \times \mathbf{n})_\Gamma = \mathbf{0}\}$; the sesquilinear forms $a_L(\cdot,\cdot) : V_L \times V_L \to \mathbb{C}$ and the linear functionals $F_L : V_L \to \mathbb{C}$ such that $a(\mathbf{v},\mathbf{w}) = a_C(\mathbf{v}_C,\mathbf{w}_C) + a_I(\mathbf{v}_I,\mathbf{w}_I)$ and $F(\mathbf{v}) = F_C(\mathbf{v}_C) + F_I(\mathbf{v}_I) \quad \forall \mathbf{v},\ \mathbf{w} \in V$. If *the sesquilinear forms $a_L(\cdot,\cdot)$ are continuous and coercive in $V_{L,0}$ for both $L = C, I$* we can define the extension operators $\mathbf{R}_L : X \to V_L$ in the following way: for any $\boldsymbol{\eta} \in X$, $\mathbf{R}_L\boldsymbol{\eta}$ is the unique function in V_L such that

$$(\mathbf{n} \times \mathbf{R}_L\boldsymbol{\eta} \times \mathbf{n})_{|\Gamma} = \boldsymbol{\eta}$$
$$a_L(\mathbf{R}_L\boldsymbol{\eta}, \mathbf{v}_L) = 0 \quad \forall \mathbf{v}_L \in V_{L,0}\,.$$

Let us consider the Steklov-Poincare operators $S_L : X \to X'$ given by

$$\langle S_L\boldsymbol{\eta}, \boldsymbol{\nu}\rangle_\Gamma = a_L(\mathbf{R}_L\boldsymbol{\eta}, \mathbf{R}_L\boldsymbol{\nu}) \quad \forall \boldsymbol{\eta},\ \boldsymbol{\nu} \in X\,.$$

Moreover we can define the functions $\hat{\mathbf{u}}_L \in V_{L,0}$ such that

$$a_L(\hat{\mathbf{u}}_L, \mathbf{v}_L) = F_L(\mathbf{v}_L) \quad \forall \mathbf{v}_L \in V_{L,0}$$

and $\boldsymbol{\chi}_L \in X'$ given by $\langle \boldsymbol{\chi}_L, \boldsymbol{\eta}\rangle_\Gamma = F_L(\mathbf{R}_L\boldsymbol{\eta}) - a_L(\hat{\mathbf{u}}_L, \mathbf{R}_L\boldsymbol{\eta}) \quad \forall \boldsymbol{\eta} \in X$. Let us denote $\boldsymbol{\chi} = \boldsymbol{\chi}_I + \boldsymbol{\chi}_C$. The Steklov-Poincare equation reads: find $\boldsymbol{\lambda} \in X$ such that

$$(S_I + S_C)\boldsymbol{\lambda} = \boldsymbol{\chi}\,. \tag{6}$$

If $\boldsymbol{\lambda}$ is solution of (6) then $\mathbf{u} = \begin{cases} \mathbf{R}_C\boldsymbol{\lambda} + \hat{\mathbf{u}}_C & \text{in } \Omega_C \\ \mathbf{R}_I\boldsymbol{\lambda} + \hat{\mathbf{u}}_I & \text{in } \Omega_I \end{cases}$ is solution of (5).

If *for one of the two subdomains the sesquilinear form $a_L(\cdot,\cdot)$ is also continuous and coercive in V_L* then for each $\boldsymbol{\xi} \in X'$ there exist a unique $\mathbf{F}_L\boldsymbol{\xi} \in V_L$ such that $a_L(\mathbf{F}_L\boldsymbol{\xi}, \mathbf{w}_L) = \langle \boldsymbol{\xi}, \mathbf{n} \times \mathbf{w}_L \times \mathbf{n}\rangle_\Gamma \quad \forall \mathbf{w}_L \in V_L$. It is easy to see that $\langle S_L(\mathbf{n} \times \mathbf{F}_L\boldsymbol{\xi} \times \mathbf{n}), \boldsymbol{\eta}\rangle_\Gamma = \langle \boldsymbol{\xi}, \boldsymbol{\eta}\rangle_\Gamma$ for all $\boldsymbol{\eta} \in X$ hence $S_L^{-1}(\boldsymbol{\xi}) = \mathbf{n} \times \mathbf{F}_L\boldsymbol{\xi} \times \mathbf{n}$. It is well known that the Dirichlet-Neumann iterative method is equivalent to the preconditioned Richardson method for the Steklov-Poincare equation

$$\boldsymbol{\lambda}^{(n+1)} = \boldsymbol{\lambda}^{(n)} + \theta S_L^{-1}\left[\boldsymbol{\chi} - (S_I + S_C)\boldsymbol{\lambda}^{(n)}\right].$$

In the **H**-based formulation the preconditioner is S_I while in the **E**-based formulation the preconditioner is S_C.

We are interested in the finite element approximation of these problems using the Nédélec curl-conforming edge elements of degree k, $N^k_{L,h} \subset H(\mathrm{curl}\,;\Omega_L)$ (see [7]) for $L = C, I$. Let us denote $\P_k$, $k \geq 0$, the space of polynomials of degree less than or equal k in the three variables x_1, x_2, x_3, and by $\widetilde{\P}_k$ the space of homogeneous polynomials of degree k. For $k \geq 1$ we define the polynomial spaces $M_k := \{\mathbf{q} \in (\widetilde{\P}_k)^3 \,|\, \mathbf{q}(\mathbf{x}) \cdot \mathbf{x} = 0\}$ and $R_k := (\P_{k-1})^3 \oplus M_k$. Let us consider a tetrahedral triangulation of Ω, $\mathscr{T}_h$, such that its restriction to Ω_L, $\mathscr{T}_{L,h}$, induces a triangulation of Ω_L. Then

$$N_{L,h} := \{\mathbf{w}_h \in H(\mathrm{curl}\,;\Omega_L) \,|\, \mathbf{w}_{h|K} \in R_k \quad \forall K \in \mathscr{T}_{L,h}\}\,.$$

We want to show that in the discrete setting the iterative procedure converges and that the convergence rate is independent of h.

The discrete $\mathbf{H}$-based formulation is stated in the space

$$V_h(\mathbf{0}) := \{\mathbf{v}_h \in N_h^k : \mathbf{v}_{I,h} \in H^0_{0,\partial\Omega}(\mathrm{curl}\,;\Omega_I)\} \subset V(\mathbf{0})\,.$$

The space X for the Dirichlet-Neumann procedure is

$$\chi_h^0 = \{(\mathbf{n}\times\mathbf{v}_h\times\mathbf{n})_{|\Gamma} : \mathbf{v}_h \in V_h(\mathbf{0})\} \subset H_0^{-1/2}(\mathrm{curl}_{\,\tau};\Gamma)\,.$$

In Ω_C we use the standard Nédélec finite elements $N^k_{C,h}$, while in Ω_I we have the finite element space

$$V_{I,h}(\mathbf{0}) = N^k_{I,h} \cap H^0_{0,\partial\Omega}(\mathrm{curl}\,;\Omega_I)\,.$$

Remark 3. Let $L^k_{I,h} \subset H^1(\Omega_I)$ be the space of standard Lagrange finite elements of degree k and $H_{I,h,0} = L^k_{I,h} \cap H^1_{0,\partial\Omega}(\Omega_I)$. Then

$$V_{I,h}(\mathbf{0}) = \nabla H_{I,h,0} + \mathscr{H}_{I,h}$$

where $\mathscr{H}_{I,h}$ is a space whose dimension coincides with n_Γ, the first Betti number of Ω_I. More precisely, there exits a system of cutting surfaces Ξ_l, $l = 1,\ldots,n_\Gamma$ with $\partial\Xi_l \subset \Gamma$ such that every function $\mathbf{v}_I \in H_{0,\partial\Omega}(\mathrm{curl}\,;\Omega_I)$ restricted to $\Omega_I \setminus \cup_{l=1}^{n_\Gamma}\Xi_l$ is the gradient of a function belonging to $H^1(\Omega_I \setminus \cup_{l=1}^{n_\Gamma}\Xi_l)$ (see e.g. [3, 5, 6]). If the triangulation $\mathscr{T}_{I,h}$ induces a triangulation on each surface Ξ_l the space $\mathscr{H}_{I,h}$ is the one generated by the $(L^2(\Omega_I))^3$-extension of the gradient of the piecewise linear function taking value one at the node on one side of Ξ_l and value zero at all the other nodes including those on the other side of Ξ_l (see [2], Sect. 5.4).

Concerning the $\mathbf{E}$-based formulation, for its finite element approximation we consider the space

$$W_h := \{\mathbf{w}_h \in N_h^k : \int_{\Omega_I} \mathbf{w}_h \cdot \boldsymbol{\nabla}\overline{\phi}_{I,h} = 0 \quad \forall\,\phi_{I,h} \in H^k_{I,h,*}\}$$

where $H^k_{I,h,*} = L^k_{I,h} \cap H^1_*(\Omega_I)$. (Notice that W_h is not a subspace of W.) The space X where the Steklov-Poincare operators are defined is the space of discrete traces

$$\chi_h = \{(\mathbf{n}\times\mathbf{w}_h\times\mathbf{n})_{|\Gamma} : \mathbf{w}_h \in N_h^k\} \subset H^{-1/2}(\mathrm{curl}_{\,\tau};\Gamma)\,.$$

Also in this case we use the standard Nédélec finite elements $N^k_{C,h}$ in Ω_C while in Ω_I we consider the finite element space

$$W_{I,h} := \{\mathbf{w}_{I,h} \in N^k_{I,h} : \int_{\Omega_I} \mathbf{w}_{I,h} \cdot \nabla\overline{\phi}_{I,h} = 0 \quad \forall\,\phi_{I,h} \in H^k_{I,h,*}\}\,.$$

In order to prove the convergence of the iterative procedure let us proceed as in [1]. If $k \in \mathbb{C}$ is an eigenvalue of the map $T_L : X \to X$, $T_L\boldsymbol{\eta} := \boldsymbol{\eta} - \theta S_L^{-1}(S_I + S_C)\boldsymbol{\eta}$

with $L = I$ or $L = C$, then $k = 1 - \theta \frac{\langle (S_I + S_C)\boldsymbol{\eta}, \boldsymbol{\eta} \rangle_\Gamma}{\langle S_L \boldsymbol{\eta}, \boldsymbol{\eta} \rangle_\Gamma} = (1-\theta) - \theta \frac{\langle S_M \boldsymbol{\eta}, \boldsymbol{\eta} \rangle_\Gamma}{\langle S_L \boldsymbol{\eta}, \boldsymbol{\eta} \rangle_\Gamma}$ for any eigenvector $\boldsymbol{\eta} \in X$. Here $M = I$ or $M = C$ but $M \neq L$. If

$$\mathrm{Re}[\langle S_I \boldsymbol{\eta}, \boldsymbol{\eta} \rangle_\Gamma] \mathrm{Re}[\langle S_C \boldsymbol{\eta}, \boldsymbol{\eta} \rangle_\Gamma] + \mathrm{Im}[\langle S_I \boldsymbol{\eta}, \boldsymbol{\eta} \rangle_\Gamma] \mathrm{Im}[\langle S_C \boldsymbol{\eta}, \boldsymbol{\eta} \rangle_\Gamma] \geq 0 \tag{7}$$

and $0 \leq \theta \leq 1$ then

$$|k|^2 \leq (1-\theta)^2 + \theta^2 \frac{|\langle S_M \boldsymbol{\eta}, \boldsymbol{\eta} \rangle_\Gamma|^2}{|\langle S_L \boldsymbol{\eta}, \boldsymbol{\eta} \rangle_\Gamma|^2} \leq (1-\theta)^2 + \theta^2 \frac{\beta_M^2}{\alpha_L^2}$$

being β_M the continuity constant of S_M and α_L the coercivity constant of S_L. Choosing $0 < \theta < \min\left(1, \frac{2\alpha_L^2}{\alpha_L^2 + \beta_M^2}\right)$ on has $|k| < 1$ for each k eigenvalue of T, hence in the discrete setting the Dirichlet-Neumann procedures converges and, if α_L and β_M are independent of the mesh size, h, also the convergence rate is independent of h.

In the **H**-based formulation we have $L = I$ and $M = C$. The sesquilinear form

$$a_C(\mathbf{v}_C, \mathbf{w}_C) := \int_{\Omega_C} \left(-i\omega^{-1} \sigma^{-1} \mathrm{curl}\, \mathbf{v}_C \cdot \mathrm{curl}\, \overline{\mathbf{w}}_C + \mu \mathbf{v}_C \cdot \overline{\mathbf{w}}_C \right)$$

is clearly continuous and coercive in $H(\mathrm{curl}\,; \Omega_C)$ hence in $N_{C,h}^k$. In the insulator $a_I(\mathbf{v}_I, \mathbf{w}_I) := \int_{\Omega_I} \mu \mathbf{v}_I \cdot \overline{\mathbf{w}}_I$ is continuous and coercive in $H^0(\mathrm{curl}\,; \Omega_I)$ then also in $V_{I,h}^0$. The coercivity of S_I with a constant α_I independent of h follows from the coercivity of $a_I(\cdot,\cdot)$ and the continuity of the trace operator while the continuity of S_C with a constant β_C independent of h follows from the continuity of $a_C(\cdot,\cdot)$ and the existence of a continuous extension operator $\mathscr{E}_{C,h} : \chi_h \to N_{C,h}^k$ with continuity constant independent of h. Such an extension has been constructed in [1]. Moreover (7) clearly holds because it reduces to $\left(\int_{\Omega_C} \mu \mathbf{R}_C \boldsymbol{\eta} \cdot \overline{\mathbf{R}_C \boldsymbol{\eta}} \right) \left(\int_{\Omega_I} \mu \mathbf{R}_I \boldsymbol{\eta} \cdot \overline{\mathbf{R}_I \boldsymbol{\eta}} \right) \geq 0$. Hence taking θ small enough the iterative Dirichlet-Neumann procedure for the **H**-based formulation converges with a rate independent of the mesh size.

On the other hand for the **E**-based formulation we have $L = C$ and $M = I$. Again the sesquilinear form

$$a_C(\mathbf{v}_C, \mathbf{w}_C) := \int_{\Omega_C} \left(\mu^{-1} \mathrm{curl}\, \mathbf{v}_C \cdot \mathrm{curl}\, \overline{\mathbf{w}}_C + i\omega\sigma \mathbf{v}_C \cdot \overline{\mathbf{w}}_C \right)$$

is clearly continuous and coercive in $H(\mathrm{curl}\,; \Omega_C)$ hence in $N_{C,h}^k$. The coercivity of S_C (the preconditioner in this case) with a constant α_C independent of h follows from the uniform coercivity of $a_C(\cdot,\cdot)$ and the continuity of the trace operator. In the insulator we have $a_I(\mathbf{v}_I, \mathbf{w}_I) := \int_{\Omega_I} \mu^{-1} \mathrm{curl}\, \mathbf{v}_I \cdot \mathrm{curl}\, \overline{\mathbf{w}}_I$ that is continuous in $H(\mathrm{curl}\,; \Omega_I)$, hence in $W_{I,h}$. Proceeding as in [2], Sect. 5.5, it can be proved that it is coercive in $W_{I,h} \cap H_0(\mathrm{curl}\,; \Omega_I)$. In order to prove the continuity of S_I with a constant β_I independent of h we need a continuous extension operator $\mathscr{E}_{I,h} : \chi_h \to W_{I,h} \cap H_{0,\partial\Omega}(\mathrm{curl}\,; \Omega_I)$. We know that there exists a continuous extension $\widehat{\mathscr{E}}_{I,h} : \chi_h \to N_{I,h}^k \cap H_{0,\partial\Omega}(\mathrm{curl}\,; \Omega_I)$ (see again [1]). Given $\boldsymbol{\eta}_h \in \chi_h$ let $\Phi_{I,h} \in H_{I,h,*}^k$ be such that

$$\int_{\Omega_I} \nabla \Phi_{I,h} \cdot \nabla \psi_{I,h} = \int_{\Omega_I} \widehat{\mathscr{E}}_{I,h} \boldsymbol{\eta}_h \cdot \nabla \psi_{I,h} \quad \forall \psi_{I,h} \in H^k_{I,h,*}.$$

Then $\mathscr{E}_{I,h}\boldsymbol{\eta}_h := \widehat{\mathscr{E}}_{I,h}\boldsymbol{\eta}_h - \nabla \Phi_{I,h}$ is a continuous extension from χ_h in the space $W_{I,h} \cap H_{0,\partial\Omega}(\text{curl}\,;\Omega_I)$ with continuity constant independent of h. Condition (7) reduce in this case to $\left(\int_{\Omega_C} \mu^{-1} \text{curl}\, \mathbf{R}_C\boldsymbol{\eta} \cdot \text{curl}\, \overline{\mathbf{R}_C\boldsymbol{\eta}}\right)\left(\int_{\Omega_I} \mu^{-1} \text{curl}\, \mathbf{R}_I\boldsymbol{\eta} \cdot \text{curl}\, \overline{\mathbf{R}_I\boldsymbol{\eta}}\right) \geq 0$ that clearly holds true.

4 Conclusion

We proposed two iterative substructuring methods for two different formulations of the eddy current problem based on the electric field and magnetic field, respectively, and provided the convergence analysis. Both formulations use a constrained space in the insulator. In the **E**-based formulation the constrain is imposed introducing a Lagrange multiplier while in the **H**-based formulation a finite element approximation $V_{I,h}(\mathbf{0})$ of the constrained space $H_{0,\partial\Omega}(\text{curl}\,;\Omega_I)$ is used. The dimension of $V_{I,h}(\mathbf{0})$ is equal to n_Γ, the dimension of the $\mathscr{H}_{I,h}$, plus the dimension of $H_{I,h,0}$, that is a space of scalar functions. So the subproblem in the insulator is smaller for the **H**-based formulation than for the **E**-based formulation. However the construction of a base of $\mathscr{H}_{I,h}$ requires the determination of a system of cutting surfaces. This procedure can be cumbersome in complex geometry configurations (for instance if the conductor is a trefoil knot) an the **E** based formulation avoids this difficult.

Bibliography

[1] A. Alonso and A. Valli. An optimal domain decomposition preconditioner for low-frequency time-harmonic Maxwell equations. *Math. Comp.*, 68(226):607–631, April 1999.

[2] A. Alonso Rodríguez and A. Valli. *Eddy Current Approximation of Maxwell Equations*, volume 4 of *Modeling, Simulation and Applications*. Springer - Verlag, Italia, 2010.

[3] A. Bossavit. *Computational Electromagnetism. Variational Formulation, Complementarity, Edge Elements*. Academic Press, San Diego, 1998.

[4] A. Buffa, M. Costabel, and D. Sheen. On traces for $\mathbf{H}(\mathbf{curl}, \Omega)$ in Lipschitz domains. *J. Math. Anal. Appl.*, 276(2):845–867, 2002.

[5] P.W. Gross and P.R. Kotiuga. Finite element-based algorithms to make cuts for magnetic scalar potentials: topological constraints and computational complexity. In F.L. Teixeira, editor, *Geometric Methods for Computational Electromagnetics*, pages 207–245. EMW Publishing, Cambridge, MA, 2001.

[6] R. Hiptmair. Finite elements in computational electromagnetism. *Acta Numerica*, pages 237–339, 2002.

[7] J.C. Nédélec. Mixed finite elements in R^3. *Numer. Math.*, 35:315–341, 1980.

[8] A. Quarteroni and A. Valli. *Domain Decomposition Methods for Partial Differential Equations*. Oxford University Press, 1999.

Mesh Regularization in Bank-Holst Parallel *hp*-Adaptive Meshing

Randolph E. Bank *[1] and Hieu Nguyen†[2]

[1] Department of Mathematics, University of California, San Diego, La Jolla, California 92093-0112, rbank@ucsd.edu.

[2] Department of Computer Science, University of California, Davis, Davis, California 95616, htrnguyen@ucdavis.edu.

1 Introduction

In this work, we study mesh regularization in Bank-Holst parallel adaptive paradigm when adaptive enrichment in both h (geometry) and p (degree) is used. The paradigm was first introduced by Bank and Holst in [1–3] and later extended to hp-adaptivity in [5]. In short, the paradigm can be summarized in the following steps.

Step 1 – Load Balancing: The problem is solved on a coarse mesh, and available a posteriori error estimates are used to partition the mesh into subregions. The partition is such that each subregion has approximately the same error although subregions may vary considerably in terms of number of elements, number of degrees of freedom, and polynomial degree.

Step 2 – Adaptive Meshing: Each processor is provided with complete data for the coarse problem and instructed to sequentially solve the *entire* problem, with the stipulation that its adaptive enrichment (in h or p) should be limited largely to its own subregion. The target number of degrees of freedom for each processor is the same.

Step 3 – Mesh Regularization: The local mesh on each processor is regularized such that the mesh for the global problem described in Step 4 is conforming in both h and p.

Step 4 – Global Solve: The final global problem consists of the union of the refined partitions provided by each processor. A final solution is computed using domain decomposition.

This paradigm is attractive as it requires low communication and allows existing sequential adaptive finite element codes to run in parallel environment without much

* The work of this author was supported by the U.S. National Science Foundation under contract DMS-0915220. The Beowulf cluster used for the numerical experiments was funded by NSF SCREMS-0619173.

† The work of this author was supported by the National Science Foundation under contract DMS-0915220 and a grant from the Vietnam Education Foundation (VEF).

R. Bank et al. (eds.), *Domain Decomposition Methods in Science and Engineering XX*, Lecture Notes in Computational Science and Engineering 91,
DOI 10.1007/978-3-642-35275-1_10,

effort in recoding. However, it also poses some challenges in mesh regularization (Step 3). Since the adaptive enrichment on each processor (Step 2) is completely independent of what happens on other processors, the global refined mesh, constructed from the meshes associated with the refined regions on each of the processors, is initially non-conforming along the interface system.[3] Thus, we need to efficiently identify and resolve these nonconformities, and ultimately to establish links between degrees of freedom on the fine mesh interface system on a given processor and the corresponding degrees of freedom on the other processors which share its interface. These tasks are challenging due to the fact that the meshes are unstructured in geometry (in h), have variable degree (variable p), no element refinement tree is available, and nonconformity exists in both h and p.

2 Data Structures

In our implementation of Bank-Holst paradigm in PLTMG, a relaxed version of longest edge bisection h-refinement and a rather flexible p-refinement strategy are used for hp-refinement, see [7].

2.1 Boundary Edge Data Structure

Each boundary edge is represented by a column in the $6 \times NBF$ integer array IBNDRY, where NBF is the number of boundary edges. For the Ith column of IBNDRY (see Table 1), four of the six entries contain information about the endpoint vertices, and indication of whether the edges is curved or straight, and a user-defined label. One entry indicates edge type (various boundary condition types, or internal interface), and the fifth entry, nonzero only for edges defining the interface system used in the parallel computation, encodes information which is used in the regularization process. This entry is described in more detail in Sect. 2.2.

Table 1. Boundary edge information

IBNDRY(1,I)	First vertex number
IBNDRY(2,I)	Second vertex number
IBNDRY(3,I)	Curved edge
IBNDRY(4,I)	Edge type
IBNDRY(5,I)	Parallel information
IBNDRY(6,I)	User label

[3] The term "interface" is used to refer to the system of edges that are shared by two subregions, and the term "boundary" is used to refer to the union of the physical boundary of the domain and the interface.

2.2 Interface Edge Labeling

One approach to solve the nonconformities in the global refined mesh is to build and store refinement trees for all elements. However, such trees lose some of their attractiveness if procedures such as mesh moving and edge flipping destroy their properties. In addition, we only need the information about the edges on the interface system, which typically is a very small fraction of the total information describing the mesh. Thus, instead of creating refinement trees for all elements, during the regularization phase we recover a refinement tree for each interface edge that defines the initial interface system. To insure that subregions remain geometrically conforming on all processors, we forbid mesh moving and edge flipping for all vertices and edges lying on the interface system.

Only minimal information needed to recover the edge refinement tree is stored for each interface edge. In particular, for each interface edge E, we need the index of its original edge $r(E)$ in the interface system of the broadcast coarse mesh (after Step 1) and its position in the refinement binary tree $s(E)$. Because the original (interface) edges are the same on all processors, we can first match them, and then their descendants based on their positions in the refinement trees. These two pieces of information are combined to make a single integer, $label(E)$, the parallel information for edge E stored in the fifth row of the IBNDRY array:

$$label(E) = r(E) + (s(E) - 1) * base.$$

Here *base* is an integer which is larger than the number of boundary edges *NBF* in the broadcast coarse mesh. For edge E_{org} in the broadcast mesh, $r(E_{org})$ is its number in the IBNDRY system and $s(E_{org}) = 1$. When an edge E is refined into two children E_1 and E_2, their labels are determined from $label(E)$ and the following identities:

$$\begin{aligned} r(E_1) &= r(E_2) = r(E) \\ s(E_1) &= 2 * s(E) \\ s(E_2) &= 2 * s(E) + 1 \end{aligned}$$

For consistency, E_1 and E_2 are ordered in the counterclockwise traversal defined by vertices of E.

2.3 Interface Data Structure

When a boundary edge is refined, its entries in IBNDRY are replaced by those of one of its children. Thus IBNDRY contains only refined boundary edges. To recover the refinement trees of the interface edges, first all of the refined edges are sorted in groups according to $r(E)$. The refined edges in each group are then ordered in a counterclockwise traversal of the interface based on their vertices (end points). Edges in each group will be used to recover a refinement tree whose leaves and root represent themselves and their original edge respectively.

In order to illustrate the construction of the refinement tree of edges sharing the same ancestor, we consider the group of all refined edges associated with the original

edge E as shown in Fig. 1. These edges have the same index $r(E)$ and have been ordered via a counterclockwise traversal. For simplicity, only positions of these edges in the binary tree are shown. First, leaf nodes for the refined edges are created. Since the two nodes with largest keys (nodes 15 and 14 in our example) are siblings, their $s(E)$ values are used to create the node of their parent (node 7). Then the parent node for the two nodes with the next largest keys (nodes 10 and 11 in our example) are created and so on. The process is completed when the root node (with key 1) is created.

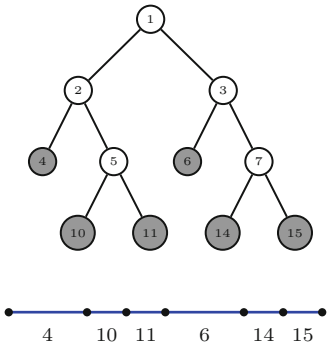


Fig. 1. Refinement tree associated with an original edge

Following the above procedure, we construct the interface data IPATH outlined in Table 2. Each interface edge, including those associated with internal nodes in refinement trees, is represented by a column with six entries in IPATH array. The first entry contains the index $r(E)$ if the edge is original (root) and zero otherwise. When edges from the two sides of the interface are matched, this entry is updated with the index of the corresponding edge. The second entry stores either the index of the edge's first child or its number in IBNDRY array (with minus sign) if it has no child. Sibling edges are put consecutively in IPATH array so storing the index for the second child edge is not necessary. Depending on the stage in the construction of IPATH array, the third and forth entries accommodate the indices of either edges, vertices or degrees of freedom of the two ends of the edge. The fifth entry is either the first or last (with minus sign) index of the interior degree(s) of freedom of the edge. This information together with the degree of the edge stored in the last entry are sufficient to recover all indices of the edge's interior degrees of freedom as they are numbered consecutively. The sign of the fifth entry indicates if they are increase or decrease along the counterclockwise traversal of the interface.

Table 2. Interface data structure: tree section

tree section				
type	root	root/leaf	internal	leaf
IPATH(1,*)	-l/n	-l/n	0/n	0/n
IPATH(2,*)	child	-e	child	-e
IPATH(3,*)	e1/v1/d1	v1/d1	e1/v1/d1	v1/d1
IPATH(4,*)	e2/v2/d2	v2/d2	e2/v2/d2	v2/d2
IPATH(5,*)	+-d	+-d	+-d	+-d
IPATH(6,*)	degree	degree	degree	degree
l=label, n=neighbor, e = edge k, v = vertex, d = dof				

3 Mesh Regularization

The regularization phase requires two all-to-all communication steps. The first describes the initial (non-conforming in h and p) interface system, and the second describes the final conforming system.

3.1 Data Reordering

At the beginning of the regularization step, each processor reorders its data structures. For processor I, edges, vertices and degrees of freedom on the interface between subregion I and the rest of the domain (fine interface) appear first in their respective arrays. These data are also arranged in a counterclockwise traversal of the interface to aid in the creation of the parallel interface data structure IPATH. Next, in all arrays, appears data corresponding to the interior of subregion I (fine interior); typically this is the majority of the data on processor I. Then appears data corresponding to the coarse part of the interface system on processor I (the interface not bounding region I). Finally appears data corresponding to the interiors of subregions other than I. Note that the first two blocks of this data (fine interface and fine interior) represent the contribution of processor I to the global fine mesh.

The parallel interface data structure IPATH is arranged in two sections; at the beginning is a pointer section with pointers for each processor's contribution to the fine interface system, and then two special sets of pointers, one for the local coarse interface system and one for the global fine mesh as a whole (see Table 3). The second section contains the tree data for individual edges on the interface system. After regularization, each processor has an IPATH array that contains complete data of the two-sided global fine interface system appended with data of local coarse interface system.

3.2 Fine Mesh Regularization

After reordering and a global exchange of interface data, each processor has complete information of the fine interface system. Then each process matches its interface edges against those of it neighbors. First original coarse edges are matched

Table 3. Interface data structure: pointer section

pointer section: 1→p+2	
IPATH(1,I)	first interface tree entry for subregion I
IPATH(2,I)	last interface tree entry for subregion I
IPATH(3,I)	first interface vertex/dof for subregion I
IPATH(4,I)	last interface vertex/dof for subregion I
$I = p + 1$: pointers for local coarse system	
$I = p + 2$: pointers for global fine system	

based on their labels. Then their descendants are matched following the refinement tree structures. We note here that for two neighboring processors, counterclockwise traversals of the interface are in opposite directions. An example of descendants of two original edges (from two different processors) is shown in Fig. 2.

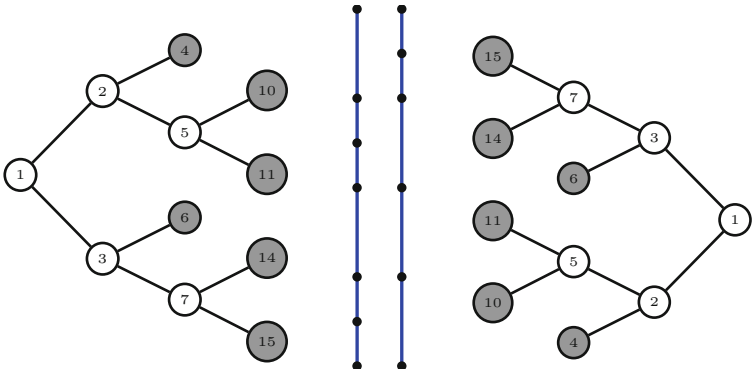


Fig. 2. Edge matching

When a pair of matching edges is determined, their first entries in IPATH are updated to store the indices (also in IPATH array) of their neighbors (change status from "−l" or "0" to "n" as in Table 2). If edges without corresponding neighbors are found, this indicates nonconformity in h. This is resolved by the processor with the less refined interface; it executes appropriate steps of h-refinement to make its interface match that of it neighbor. Although we must allow for arbitrary differences in refinement, it is typical to see at most one level of refinement difference on the fine portion of the interface. An example in Fig. 2 is edge 4 on the left that corresponds to edge 7 on the right with two child edges 14 and 15. In this case, edge 4 on the left will be h-refined one level.

When issues of h-conformity are resolved, the edges are re-examined to eliminate nonconformity in degree. Since the mesh is now h-conforming, each leaf edge on the fine interface system should have exactly one matching neighbor (from an-

other processor). If the degrees of a matching pair are different, this nonconformity is resolved by the processor with the edge of lower degree; it executes appropriate steps of p-refinement in order to achieve the same degree as its neighbor on the interface edge. However, if red-green like refinement rules are applied as in [6], fixing the degree for one interface edge might also change the degree of another interface edge and cause further nonconformity. Thus, multiple communication steps might be required to eliminate nonconformity in degree. This issue was the main motivation for us to find a more flexible p-refinement algorithm and more general nodal basis functions for transition elements, allowing the mesh to be made both h and p conforming with just one communication step. Such approach is described in [5–7].

When the global mesh is conforming, a second reordering as described above is carried out locally on each processor, followed by a second all-to-all broadcast of the new IPATH array. This time no nonconforming edges will be encountered during the matching process.

3.3 Coarse Mesh Regularization

The coarse part of the local mesh on processor I allows a complete conforming mesh of the whole domain on each processor, thus avoiding otherwise necessary communication steps. Due to constraints of shape regularity, the coarse mesh will typically be reasonably fine in areas near the fine subregion Ω_I and become more coarse in regions more distant from Ω_I. However, in some special situations such as having a singularity outside of Ω_I, the coarse mesh on processor I might be refined [8]. In very unusual circumstances, it is possible for the coarse mesh on some processors to be more refined (in h or in p) than the global fine mesh in some areas. Although this does not influence the global fine mesh solution directly, our DD solver assumes that the coarse mesh on each processor is not more refined than the global fine mesh, see [4, 9].

As described in Sect. 3.1, the IPATH array on each processor has a section for the coarse interface edges; this part of the data structure is local and different on every processor. Following the second and final broadcast of the IPATH data structure, each coarse interface edge is matched with one of the global fine edges. Here, the matching is one-way from a coarse edge to a fine edge only. Based on this type of matching, over-refined coarse edges are identified and then unrefined in either h or p.

We have also observed empirically [5, 9] that the convergence properties of our DD solver are enhanced when elements in the coarse regions having edges on the coarse interface system are more refined than those in the interior parts of the coarse region. To capture this effect, we also allow some limited refinement of elements lying along the coarse interface. The level of refinement on the interface boundary of Ω_J is determined by its distance from Ω_I; distance is measured in a graph in which the Ω_J correspond to vertices and the edge between Ω_I and Ω_J is present if and only if they have a shared interface boundary. The level of allowed refinement decays as 2^{-K}, where K is the distance from Ω_I to Ω_J.

Bibliography

[1] Randolph E. Bank. Some variants of the Bank-Holst parallel adaptive meshing paradigm. *Comput. Vis. Sci.*, 9(3):133–144, 2006.

[2] Randolph E. Bank and Michael Holst. A new paradigm for parallel adaptive meshing algorithms. *SIAM J. Sci. Comput.*, 22(4):1411–1443 (electronic), 2000.

[3] Randolph E. Bank and Michael Holst. A new paradigm for parallel adaptive meshing algorithms. *SIAM Rev.*, 45(2):291–323 (electronic), 2003. Reprinted from SIAM J. Sci. Comput. **2**2 (2000), no. 4, 1411–1443 [MR1797889].

[4] Randolph E. Bank and Shaoying Lu. A domain decomposition solver for a parallel adaptive meshing paradigm. *SIAM J. Sci. Comput.*, 26(1):105–127 (electronic), 2004.

[5] Randolph E. Bank and Hieu Nguyen. Domain decomposition and *hp*-adaptive finite elements. In Yunqing Huang, Ralf Kornhuber, Olof Widlund, and Jinchao Xu, editors, *Domain Decomposition Methods in Science and Engineering XIX*, Lecture Notes in Computational Science and Engineering, pages 3–13, Berlin, 2011. Springer-Verlag.

[6] Hieu Nguyen. *p*-adaptive and automatic *hp*-adaptive finite element methods for elliptic partial differential equations Ph.D. Thesis, Department of Mathematics, University of California, San Diego, La Jolla, CA, 2010.

[7] Randolph E. Bank and Hieu Nguyen. *hp* adaptive finite elements based on derivative recovery and superconvergence. *Comput. Vis. Sci.*, 14(6):287–299, 2011.

[8] Randolph E. Bank and Jeffrey S. Ovall. Dual functions for a parallel adaptive method. *SIAM J. Sci. Comput.*, 29(4):1511–1524 (electronic), 2007.

[9] Randolph E. Bank and Panayot S. Vassilevski. Convergence analysis of a domain decomposition paradigm. *Comput. Vis. Sci.*, 11(4–6):333–350, 2008.

Robust Parameter-Free Multilevel Methods for Neumann Boundary Control Problems

Etereldes Gonçalves[1] and Marcus Sarkis[2]

[1] Universidade Federal do Espírito Santo - UFES - etereldes@gmail.com
[2] Mathematical Sciences Department, Worcester Polytechnic Institute, USA and Instituto de Matemática Pura e Aplicada - IMPA, Brazil msarkis@wpi.edu

Summary. We consider a linear-quadratic elliptic control problem (LQECP). For the problem we consider here, the control variable corresponds to the Neumann data on the boundary of a convex polygonal domain. The optimal control unknown is the one for which the harmonic extension approximates best a specified target in the interior of the domain. We propose a multilevel preconditioner for the reduced Hessian resulting from the application of the Schur complement method to the discrete LQECP. In order to derive robust stabilization parameters-free preconditioners, we first show that the Schur complement matrix is associated to a linear combination of negative Sobolev norms and then propose preconditioner based on multilevel methods. We also present numerical experiments which agree with the theoretical results.

1 Introduction

The problem of solving linear systems is central in numerical analysis. Systems arising from the discretization of PDEs and control problems have received special attention since they appear in many applications, such as in fluid dynamics and structural mechanics. Typically, as the dimension of the discrete space increases, the resulting system becomes very ill-conditioned. To avoid the large cost of LU factorizations of KKT saddle point linear systems, we consider instead the reduced Hessian systems. To build efficient solvers, the spectral properties of these systems must be taken into account. In this paper, we develop the mathematical tools necessary to analyze and to design solvers for a model control problem. We believe that the proposed framework can be extended to more complex control problems.

2 Setting Out the Problem

Consider the following LQECP:

$$\begin{aligned} &\text{Minimize } J(u,\lambda) := \|u-u_*\|^2_{L^2(\Omega)} + \tfrac{\alpha}{2}\,\|\lambda\|^2_{H^{-1/2}(\Gamma)} + \tfrac{\beta}{2}\,\|\lambda\|^2_{L^2(\Gamma)} \\ &\text{subject to} \qquad \begin{cases} -\Delta u(x) = f(x) & \text{in } \Omega \subset \mathbb{R}^2, \\ \gamma \frac{\partial u}{\partial \eta}(s) = -\lambda(s) & \text{on } \Gamma := \partial\Omega, \end{cases} \end{aligned} \tag{1}$$

R. Bank et al. (eds.), *Domain Decomposition Methods in Science and Engineering XX*, Lecture Notes in Computational Science and Engineering 91,
DOI 10.1007/978-3-642-35275-1_11,

where u_* and f are given functions in $L^2(\Omega)\backslash\mathbb{R}$, γ is the trace operator on Γ, and α and β are nonnegative given stabilization parameters. The minimization is taken on $u \in H^1(\Omega)\backslash\mathbb{R}$ and $\lambda \in L^2(\Gamma)\backslash\mathbb{R}$. Here, "$\backslash\mathbb{R}$" stands for functions with zero average on Ω or Γ. We assume that the domain Ω is a convex polygonal domain, hence, H^2-regularity of u is assumed. The norm $H^{-1/2}(\Gamma)$ is defined as

$$\|\lambda\|^2_{H^{-1/2}(\Gamma)} := |v_\lambda|^2_{H^1(\Omega)}, \tag{2}$$

where $v_\lambda \in H^1(\Omega)\backslash\mathbb{R}$ is the harmonic extension of λ in Ω. We remark that the assumption $\alpha+\beta>0$ is necessary for the well-posedness of the problem (1), see [7, 9, 11] and references therein. The case $\alpha=\beta=0$ can also be treated by enlarging the minimizing space for λ from $H^{-1/2}(\Gamma)\backslash\mathbb{R}$ to $H^{-3/2}_{t,00}(\Gamma)\backslash\mathbb{R}$; see [6] for details. To make the notation less cumbersome, we sometimes drop "$\backslash\mathbb{R}$" below.

We consider the following discretization for the LQECP (1). We consider the space of piecewise linear and continuous functions $V_h(\Omega)\subset H^1(\Omega)$ to approximate u and p, and $\Lambda_h(\Gamma)\subset H^{1/2}(\Gamma)$ (the restriction of $V_h(\Omega)$ to Γ) to approximate λ. The underlying triangulation $\mathcal{T}_h(\Omega)$ is assumed to be quasi-uniform with mesh size $O(h)$. Let $\{\phi_1(x),\ldots,\phi_n(x)\}$ and $\{\varphi_1(x),\ldots,\varphi_m(x)\}$ denote the standard hat nodal basis functions for $V_h(\Omega)$ and $\Lambda_h(\Gamma)$, respectively. The corresponding discrete problem associated to (1) results in

$$\begin{bmatrix} M & 0 & A^T \\ 0 & G & Q^TE^T \\ A & EQ & 0 \end{bmatrix} \begin{bmatrix} \mathrm{u} \\ \lambda \\ \mathrm{p} \end{bmatrix} = \begin{bmatrix} \mathrm{f}_1 \\ \mathrm{f}_2 \\ \mathrm{f}_3 \end{bmatrix}, \tag{3}$$

where the matrices M and A are the mass and stiffness matrices on Ω, and Q is the mass matrix on Γ. We define $Q_{ext_{ij}} = (\phi_i,\varphi_j)_{L^2(\Gamma)}$; $\phi_i \in V_h(\Omega)$ and $\varphi_j \in \Lambda_h(\Gamma)$. It is easy to see that $Q_{ext} = EQ$, where $E \in \mathbb{R}^{n\times m}$ is the trivial zero discrete extension operator defined from $\Lambda_h(\Gamma)$ to $V_h(\Omega)$. We define $G \in \mathbb{R}^{m\times m}$ as be the matrix associated to the norm $\frac{\alpha}{2}\|\cdot\|^2_{H_h^{-1/2}(\Gamma)} + \frac{\beta}{2}\|\cdot\|^2_{L^2(\Gamma)}$ on $\Lambda_h(\Gamma)$, where $\|\lambda\|_{H_h^{-1/2}(\Gamma)} := |v^h_\lambda|_{H^1(\Omega)}$ with $v^h_\lambda := A^\dagger Q_{ext}\lambda$, i.e., v^h_λ is the discrete harmonic extension version of (2) with $\lambda \in \Lambda_h(\Gamma)$. Hence, we have $G = \alpha(Q^T_{ext}A^\dagger)A\,(A^\dagger Q_{ext}) + \beta Q = Q^T(\alpha E^T A^\dagger E + \beta Q^{-1})Q$. Here and the following $A^\dagger$ is the pseudo inverse of A. The discrete forcing terms are defined by $(\mathrm{f}_1)_i = \int_\Omega u_*(x)\,\phi_i(x)\,dx$, for $1\le i\le n$, $\mathrm{f}_2 = 0$ and $(\mathrm{f}_3)_i = \int_\Omega f(x)\,\phi_i(x)\,dx$.

3 The Reduced Hessian $\mathscr{H}$

In this paper we propose and analyze preconditioners for the reduced Hessian associated to (3). Eliminating the variables u and p from Eq. (3), and denoting $S_1^\dagger := E^T A^\dagger E$ and $S_3^\dagger := E^T A^\dagger M A^\dagger E$, we obtain

$$\mathscr{H}\lambda := Q(\alpha S_1^\dagger + \beta Q^{-1} + S_3^\dagger)Q\lambda = b := Q^T_{ext}A^\dagger M A^\dagger \mathrm{f}_3 - Q^T_{ext}A^\dagger \mathrm{f}_1. \tag{4}$$

The matrix $\mathscr{H}$ is known as the Schur complement (reduced Hessian) with respect to the discrete control variable λ. We observe that the state variable u can be obtained

by solving (4) and using the third equation of (3). We note that the Reduced matrix $\mathscr{H}$ is a symmetric positive definite matrix on

$$\Lambda_h(\Gamma)\backslash_Q\mathbb{R} := \{\lambda \in \Lambda_h(\Gamma); (\lambda, 1)_{L^2(\Gamma)} = (Q\lambda, 1_m)_{\ell^2} = 0\},$$

hence, we consider the *Preconditioned Conjugate Gradient* (PCG) with a preconditioner acting on $\Lambda_h(\Gamma)\backslash_Q\mathbb{R}$. Note also that $A^\dagger$ is also symmetric positive definite matrix on

$$V_h(\Omega)\backslash_M\mathbb{R} := \{u \in V_h(\Omega); (u, 1)_{L^2(\Omega)} = (Mu, 1_n)_{\ell^2} = 0\}.$$

The main goal of this paper is to develop robust preconditioned multilevel methods for the matrix $\mathscr{H}$ such that the condition number estimates that do not depend on α and β, and depend on $\log^2(h)$.

We point out that several block preconditioners for solving systems like (3) were proposed in the past; see [1, 8, 11, 14] and references therein. These preconditioners depend heavily on the availability of a good preconditioner for the Schur complement matrix. To the best of our knowledge, no robust and mathematically sounded preconditioner was systematically carried out for the reduced Hessian (4). Most of the existing work is toward problems where the control variable is f rather than λ, and even for these cases, condition number estimates typically deteriorate when all the stabilization parameters go to zero. Related work to ours is developed in [13] where it is proposed a preconditioner for the first biharmonic problem discretized by the mixed finite element method introduced by Ciarlet and Raviart [4]. Using techniques developed in [5], Peisker transforms the discrete problem to an interface problem and a preconditioner based on FFT is proposed and analyzed. This approach can also be interpreted as a control problem like (1), however, replacing the Neumann control by a Dirichlet control. We note that Dirichlet control problems are much easier to handle and to study since in (4) the operator $S_3^\dagger$ is replaced by $S_1^\dagger$, and therefore, a multilevel method such as in [2], can be applied. An attempt to precondition the Neumann control problem via FFT was considered in [7], however, such as in Peisker's work, it holds only for special meshes where the Schur complement matrix and the mass matrix on Γ share the same set of eigenvectors.

4 Theoretical Remarks on the Reduced Hessian $\mathscr{H}$

In this section we associate the Reduced Hessian $\mathscr{H}$ to a linear combination of Sobolev norms. Here and below we use the notation $a \preceq (\succeq)\, b$ to indicate that $a \leq (\geq) Cb$, where the positive constant C depends only on the shape of Ω and $\boldsymbol{\tau}_h(\Omega)$. When $a \preceq b \preceq a$, we say $a \asymp b$.

First we observe that G is associated to the norm $\frac{\alpha}{2}\|\cdot\|^2_{H_h^{-1/2}(\Gamma)} + \frac{\beta}{2}\|\cdot\|^2_{L^2(\Gamma)}$ in $\Lambda_h(\Gamma)$. It is well known that for $\lambda \in \Lambda_h(\Gamma)\backslash_Q\mathbb{R}$ we have

$$\lambda^T Q S_1^{\dagger} Q \lambda = \|\lambda\|^2_{H_h^{-1/2}(\Gamma)} \asymp \|\lambda\|^2_{H^{-1/2}(\Gamma)}. \tag{5}$$

What is not obvious is how to associate the matrix $QS_3^{\dagger}Q$ to a Sobolev norm, and this is given in the following result (see [6]):

Theorem 1. *Let $\Omega \subset \mathbb{R}^2$ be a convex polygonal domain. Let $v_\lambda^h := A^{\dagger} Q_{ext}\,\lambda \in V_h(\Omega)\backslash_M \mathbb{R}$ be the discrete harmonic function with Neumann data $\lambda \in \Lambda_h(\Gamma)\backslash_Q \mathbb{R}$. Then,*

$$\lambda^T Q S_3^{\dagger} Q \lambda = \|v_\lambda^h\|^2_{L^2(\Omega)} \asymp \|\lambda\|^2_{H_{t,00}^{-3/2}(\Gamma)} + h^2\,\|\lambda\|^2_{H^{-1/2}(\Gamma)}. \tag{6}$$

Using these results we conclude that $\mathscr{H}$ is associated to the following linear combination of Sobolev norms

$$\lambda^T \mathscr{H} \lambda \asymp (\alpha + h^2)\,\|\lambda\|^2_{H^{-1/2}(\Gamma)} + \beta\,\|\lambda\|^2_{L^2(\Gamma)} + \|\lambda\|^2_{H_{t,00}^{-3/2}(\Gamma)}. \tag{7}$$

Remark 1. We next hint why the norm $\|\cdot\|^2_{H_{t,00}^{-3/2}(\Gamma)}$ is fundamental for this problem. Let $\{\Gamma_k\}_{1\le k\le K}$ and $\{\delta_k\}_{1\le k\le K}$ be the edges and the vertices of the polygonal Γ, respectively. Let $C^\infty_{t,00}(\Gamma_k) := \{\lambda \in C^\infty(\Gamma_k); \partial\lambda/\partial\tau_k \in C_0^\infty(\Gamma_k)\}$, where τ_k stands for the tangential unit vector on Γ_k. Define $H^2_{t,00}(\Gamma_k)$ by the closure of $C^\infty_{t,00}(\Gamma_k)$ in the $H^2(\Gamma_k)$-norm, that is,

$$H^2_{t,00}(\Gamma_k) := \{\lambda \in H^2(\Gamma_k); \frac{\partial\lambda}{\partial\tau_k}(\delta_{k-1}) = \frac{\partial\lambda}{\partial\tau_k}(\delta_k) = 0\}. \tag{8}$$

Using interpolation theory of operators and a characterization of $H^{3/2}_{t,00}(\Gamma_k)$, see [10], it is possible to show that

$$H^{3/2}_{t,00}(\Gamma_k) := \left[H^2_{t,00}(\Gamma_k), H^1(\Gamma_k)\right]_{1/2} = \left\{\lambda \in H^{3/2}(\Gamma_k); \partial\lambda/\partial\tau_k \in H^{1/2}_{00}(\Gamma_k)\right\}.$$

We define $H^{3/2}_{t,00}(\Gamma) = H^{1/2}(\Gamma) \cap \prod_{k=1}^K H^{3/2}_{t,00}(\Gamma_k)$ endowed with the norm

$$\|\lambda\|_{H^{3/2}_{t,00}(\Gamma)} := \|\lambda\|^2_{H^{1/2}(\Gamma)} + \sum_{k=1}^K \|\frac{\partial\lambda}{\partial\tau_k}\|^2_{H^{1/2}_{00}(\Gamma_k)}, \tag{9}$$

and define $H^{-3/2}_{t,00}(\Gamma) = (H^{3/2}_{t,00}(\Gamma))'$. The fundamental property of this space is that

$$\|\lambda\|_{H^{-3/2}_{t,00}(\Gamma)} \asymp \|v_\lambda\|_{L^2(\Omega)},$$

where v_λ is defined by (2); see [6].

5 Preconditioning Sobolev Norms Using Multilevel Methods

In this section, using multilevel based preconditioners, we develop spectral approximations for matrices associated to several Sobolev norms; see [2, 3, 12, 15], and references therein.

5.1 Notation and Technical Tools

From now on, we assume that the triangulation $\boldsymbol{\tau}_h$ of Γ has a *multilevel* structure. More precisely, denoting $\boldsymbol{\tau}_h$ as the restriction of $\boldsymbol{\tau}_h(\Omega)$ to Γ, we assume that the triangulation $\boldsymbol{\tau}_h$ is obtained from $(L-1)$ successive refinements of an initial coarse triangulation $\boldsymbol{\tau}_0$ with initial grid size h_0. We assume also that $h_\ell = h_{\ell-1}/2$ is the grid size on the ℓ-th triangulation $\boldsymbol{\tau}_\ell$ and associate the standard P_1 finite element space $V_\ell(\Gamma)$ generated by continuous and piecewise linear basis functions $\{\varphi_i^\ell\}_{i=1}^{m_\ell}$. Hence, we have

$$V_0(\Gamma) \subset V_1(\Gamma) \subset \cdots \subset V_L(\Gamma) := V_h(\Gamma) \subset L^2(\Gamma).$$

Let P_ℓ denote the $L^2(\Gamma)$-orthogonal projection onto $V_\ell(\Gamma)$, and let $\Delta P_\ell := (P_\ell - P_{\ell-1})$, that is, the $L^2(\Gamma)$-orthogonal projection onto $V_\ell(\Gamma) \cap V_{\ell-1}(\Gamma)^\perp$. We have that $P_0, (P_1 - P_0), \ldots, (P_L - P_{L-1})$ restricted to $V_L(\Gamma)$ are mutually L^2-orthogonal projections which satisfy:

$$I = P_0 + (P_1 - P_0) + \cdots + (P_L - P_{L-1}). \tag{10}$$

Note that $P_L = I$. The matrix form of P_ℓ restricted to $V_L(\Gamma)$ is given by

$$P_\ell = R_\ell^T Q_\ell^{-1} R_\ell Q, \tag{11}$$

where R_ℓ is the $m_\ell \times m_L$ restriction matrix, that is, the i-th row of R_ℓ is obtained by interpolating the basis function $\varphi_i^\ell \in V_\ell := V_\ell(\Gamma)$ at the nodes of the finest triangulation $\boldsymbol{\tau}_L := \boldsymbol{\tau}_h$.

It follows from [2, 12], that for $-3/2 < s < 3/2$

$$\|\mathrm{v}\|^2_{H^s(\Gamma)} \asymp \sum_{\ell=0}^{L} h_\ell^{-2s} \|(P_\ell - P_{\ell-1})\mathrm{v}\|^2_{L^2(\Gamma)}, \text{ for all } \mathrm{v} \in V_L. \tag{12}$$

This constraint for s comes from the fact that for $s \geq 3/2$ we have $V_h(\Gamma) \not\subset H^s(\Gamma)$, therefore, the equivalence deteriorates when s tends to $3/2$. Results for negative norms are obtained by duality.

We now describe how to represent the splitting $\sum_{\ell=0}^{L} \mu_\ell \|(P_\ell - P_{\ell-1})\mathrm{v}\|^2_{L^2(\Gamma)}$ into a matrix form. Let $\Delta_\ell := (P_\ell - P_{\ell-1})Q^{-1} = R_\ell^T Q_\ell^{-1} R_\ell - R_{\ell-1}^T Q_{\ell-1}^{-1} R_{\ell-1}$. Then we have

$$\Delta_k Q \Delta_\ell = \delta_{k\ell} \Delta_\ell \text{ and } \sum_{\ell=0}^{L} \mu_\ell \|(P_\ell - P_{\ell-1})\mathrm{v}\|^2_{L^2(\Gamma)} = \sum_{\ell=0}^{L} \mu_\ell \mathrm{v}^T Q (P_\ell - P_{\ell-1})\mathrm{v}, \tag{13}$$

where $P_{-1} = 0$. We observe that $Q(P_\ell - P_{\ell-1}) = Q\Delta_\ell Q$ is symmetric semi-positive definite. By (12) and (13), for all $\mathrm{v} \in V_L$ we have

$$\|\mathrm{v}\|^2_{H^{-1/2}(\Gamma)} \asymp (\sum_{\ell=0}^{L} h_\ell \Delta_\ell Q\mathrm{v}, Q\mathrm{v}). \tag{14}$$

To invert a matrix of the form $\sum_{k=0}^{L} \mu_k^{-1} \Delta_k Q$, we first assume that $\mu_k > 0$, $0 \leq k \leq L$. Then, from (10) and (13) we obtain

$$(\sum_{k=0}^{L} \mu_k^{-1} \Delta_k Q)(\sum_{\ell=0}^{L} \mu_\ell \Delta_\ell Q) = I. \tag{15}$$

5.2 Multilevel Preconditioner for the Reduced Hessian $\mathscr{H}$

In this subsection we analyze a multilevel preconditioner for Reduced Hessian $\mathscr{H}$. We first present a preconditioner for G as follows. Using (2), (14) and (15) we obtain

$$\begin{cases} S_1 & \asymp Q\sum_{\ell=0}^{L} h_\ell^{-1}\Delta_\ell Q, \\ QS_1^{\dagger}Q & \asymp Q\sum_{\ell=0}^{L} h_\ell \Delta_\ell Q. \end{cases} \tag{16}$$

The above equivalences yield simultaneous approximation for the spectral representations of $G := \beta Q + \alpha Q S_1^{\dagger} Q$ in terms of the Δ_ℓ and Q. More precisely,

$$G \asymp Q\sum_{\ell=1}^{L}(\beta + \alpha h_\ell)\Delta_\ell Q, \tag{17}$$

and using (15) and (17), the following spectral equivalency holds

$$G^{-1} \asymp \sum_{\ell=0}^{L}(\beta + \alpha h_\ell)^{-1}\Delta_\ell. \tag{18}$$

We next establish that $\sum_{\ell=0}^{L}(h_\ell^{-3})\Delta_\ell$ is a quasi-optimal preconditioner for $QS_3^{\dagger}Q$. More precisely, we have the following result (see [6]):

Theorem 2. *For all* $v_L \in \mathbf{V}_L$*, the following inequalities hold:*

$$\|v_L\|^2_{H_{t,00}^{-3/2}(\Gamma)} \preceq \sum_{\ell=1}^{L} h_\ell^3 \|\Delta P_\ell v_L\|^2_{L^2} \preceq (L+1)^2 \|v_L\|^2_{H_{t,00}^{-3/2}(\Gamma)}. \tag{19}$$

From Theorems 1 and 2 and (15), we establish the main result, the quasi-optimality for a preconditioner for $\mathscr{H}$.

Theorem 3. *Let* $\mathscr{PC} := \sum_{\ell=0}^{L}(\alpha h_\ell + \beta + h_\ell^3)^{-1}\Delta_\ell$. *Then*

$$(L+1)^{-2}\mathscr{PC} \;\preceq\; \mathscr{H}^{-1} \;\preceq\; \mathscr{PC}. \tag{20}$$

6 Numerical Results

In this section we show numerical results conforming the theory developed. For all tests presented, Ω is the square domain $[0,1]\times[0,1]$. The triangulation of Ω is constructed as follows. We divide each edge of $\partial\Omega$ into 2^N parts of equal length, where N is an integer denoting the number of refinements. In all tests (*cond*) means condition number, (*it*) indicates the number of iterations of the PCG, (*eig min*) means the lowest eigenvalue for preconditioned system. To calculate the eigenvalues we build the preconditioned system and use the function *eig* of MATLAB. We can see from tables below the asymptotic $\log^2(h)$ behavior for the case $\alpha = \beta = 0$, i.e., $\mathrm{cond}(N+1) - \mathrm{cond}(N)$ grows linearly with N. As expected, larger is α or β, better conditioned are the preconditioned systems (Tables 1–4).

Remark 2. Numerical experiments show (not reported here) that the largest eigenvalue of $\left(\sum_{\ell=0}^{L}\Delta_\ell\right) * Q$ divided by the largest eigenvalue of $\left(\sum_{\ell=0}^{L} h_\ell^{-3}\Delta_\ell\right) * QS_3^{\dagger}Q$ converges to 36 when h decreases to zero. In tables above, we considered the rescaled preconditioner

$\mathscr{PC}_r * \mathscr{H}$ with $\beta = 1$				$\mathscr{PC}_r * \mathscr{H}$ with $\beta = (0.1)^3$		
N $\downarrow$	cond	eig min	it	cond	eig min	it
4	1.04237	0.02756	2	4.94294	0.01622	7
5	1.04222	0.02757	2	4.87258	0.01655	7
6	1.04218	0.02757	2	4.85515	0.01663	7
7	1.04217	0.02757	2	4.85084	0.01665	7

Table 1. Equivalence between $\mathscr{H}$ and $\mathscr{PC}_r$ with $r = 36$ and $\alpha = 0$.

$\mathscr{PC}_r * \mathscr{H}$ with $\beta = (0.1)^6$				$\mathscr{PC}_r * \mathscr{H}$ with $\beta = 0$		
N $\downarrow$	cond	eig min	it	cond	eig min	it
4	28.1662	0.004747	15	33.5522	0.004016	16
5	24.3303	0.005739	20	41.9737	0.003407	25
6	20.3042	0.006984	22	50.5193	0.002930	35
7	18.9576	0.007514	20	59.2085	0.002550	44

Table 2. Equivalence between $\mathscr{H}$ and $\mathscr{PC}_r$ with $r = 36$ and $\alpha = 0$.

$\mathscr{PC}_r * \mathscr{H}$ with $\alpha = 1$				$\mathscr{PC}_r * \mathscr{H}$ with $\alpha = (0.1)^3$		
N $\downarrow$	cond	eig min	it	cond	eig min	it
4	4.62312	0.11893	10	13.7601	0.010698	14
5	5.12018	0.11826	10	18.3917	0.012503	19
6	5.33402	0.11798	11	26.2878	0.013139	22
7	5.45327	0.11788	12	35.6393	0.013312	26

Table 3. Equivalence between $\mathscr{H}$ and $\mathscr{PC}_r$ with $r = 36$ and $\beta = 0$.

$\mathscr{PC}_r * \mathscr{H}$ with $\alpha = (0.1)^6$			$\mathscr{PC}_r * \mathscr{H}$ with $\alpha = 0$		
4	33.4363	0.004031	16	33.5522	0.0040164	16
5	41.4318	0.003452	25	41.9737	0.0034074	25
6	48.1852	0.003073	33	50.5193	0.0029301	35
7	50.8326	0.002973	43	59.2085	0.0025501	44

Table 4. Equivalence between $\mathscr{H}$ and $\mathscr{PC}_r$ with $r = 36$ and $\beta = 0$.

$$\mathscr{PC}_r := \sum_{\ell=0}^{L} (\alpha h_\ell + r\beta + h_\ell^3)^{-1} \Delta_\ell,$$

with $r = 36$, instead of $\mathscr{PC} := \sum_{\ell=0}^{L} (\alpha h_\ell + \beta + h_\ell^3)^{-1} \Delta_\ell$. This change improves considerably the condition number of preconditioners and improve slightly the number of iterations.

Bibliography

[1] M. Benzi, G. H. Golub, and J. Liesen. Numerical solution of saddle point problems. *Acta Numerica*, pages 1–137, 2005.

[2] J. H. Bramble, J. E. Pasciak, and P. S. Vassilevski. Computational scales of Sobolev norms with application to preconditioning. *Math. Comp.*, 69:463–480, 2000.

[3] James H. Bramble, Joseph E. Pasciak, and Jinchao Xu. Parallel multilevel preconditioners. *Math. Comp.*, 55(191):1–22, 1990. ISSN 0025-5718. doi: 10.2307/2008789.

[4] P. G. Ciarlet and P.-A. Raviart. A mixed finite element method for the biharmonic equation. In *Mathematical aspects of finite elements in partial differential equations (Proc. Sympos., Math. Res. Center, Univ. Wisconsin, Madison, Wis., 1974)*, pages 125–145. Publication No. 33. Math. Res. Center, Univ. of Wisconsin-Madison, Academic Press, New York, 1974.

[5] R. Glowinski and O. Pironneau. Numerical methods for the first biharmonic equation and the two-dimensional Stokes problem. *SIAM Rev.*, 21(2):167–212, 1979.

[6] Etereldes Goncalves and Marcus Sarkis. Robust preconditioners for the Hessian system in elliptic optimal control problems. 2011.

[7] Etereldes Goncalves, Tarek P. Mathew, Marcus Sarkis, and Christian E. Schaerer. A robust preconditioner for the Hessian system in elliptic optimal control problems. In *Domain decomposition methods in science and engineering XVII*, volume 60 of *Lect. Notes Comput. Sci. Eng.*, pages 527–534. Springer, Berlin, 2008.

[8] A. Klawonn. Block triangular preconditioners for saddle point problems with a penalty term. *SIAM J. Sci. Compt.*, 19:172–184, 1998.

[9] J. L. Lions. *Some methods in the mathematical analysis of systems and their control*. Taylor and Francis, 1981.

[10] J. L. Lions and E. Magenes. *Problèmes aux limites non homogènes et applications*, volume 1. Dunod Paris, 1968.

[11] T. P. Mathew, M. Sarkis, and C. E. Schaerer. Analysis of block matrix preconditioners for elliptic optimal control problems. *Numer. Linear Algebra Appl.*, 14(4):257–279, 2007.

[12] P. Oswald. Multilevel norms for $H^{-1/2}$. *Computing*, 61:235–255, 1998.

[13] P. Peisker. On the numerical solution of the first biharmonic equation. *Mathematical Modeling and Numerical analysis*, 22:655–676, 1988.

[14] René Simon and Walter Zulehner. On Schwarz-type smoothers for saddle point problems with applications to PDE-constrained optimization problems. *Numer. Math.*, 111(3):445–468, 2009. ISSN 0029-599X. doi: 10.1007/s00211-008-0187-1.

[15] Xuejun Zhang. Multilevel Schwarz methods for the biharmonic Dirichlet problem. *SIAM J. Sci. Comput.*, 15(3):621–644, 1994.

An Overlapping Domain Decomposition Method for a 3D PEMFC Model

Cheng Wang[1], Mingyan He[1], Ziping Huang[1,2], and Pengtao Sun[3]

[1] Department of Mathematics, Tongji University, Shanghai, China
wangcheng@tongji.edu.cn hemingyan1985@yahoo.com.cn
[2] Chinese-German College, Tongji University, Shanghai, China
huangziping@tongji.edu.cn
[3] Department of Mathematical Sciences, University of Nevada Las Vegas, 4505 Maryland Parkway, Las Vegas, NV 89154 pengtao.sun@unlv.edu

Summary. In this paper, an overlapping domain decomposition method is developed to simulate the water management of the polymer exchange membrane fuel cell on the local structured grids. Numerical experiments demonstrate that our methods are effective to deal with the simulation on the non-matching grids with low mass balance error.

1 Introduction

Polymer exchange membrane fuel cells (PEMFCs) have been used in a large number of industries worldwide because of their advantages such as low environmental impact, rapid start-up and high power density [15, 16]. The performance of fuel cell is affected by many factors, such as material parameters, operating conditions, different channel structures and so on [2, 9, 10].

For better performance, different structures for the anode and cathode gas channels are used in the PEMFC practical design. This asymmetrical structure can keep the balance of pressures on both sides of the membrane. Thus the water management in cathode can be improved and the duration of fuel cell can be prolonged. An unstructured grid partitioned by tetrahedra or triangles can be used for this asymmetrical fuel cell in single domain approach, but structured grids, such as hexahedron and quadrilateral, are easily implemented and have super convergence [1, 4, 14]. However, non-matching grids would be generated when partitioning with structured grids in numerical simulations. Besides, since oxygen reduction reaction occurs in cathode, the variation of physical quantities such as water concentration are more significant in cathode than in anode. So it is necessary for cathode to simulate these phenomena accurately by a refined grid. The objective of this paper is to provide an overlapping domain decomposition method for the simulation of a 3D single-phase PEMFC model with local structured grid in anode and cathode respectively.

R. Bank et al. (eds.), *Domain Decomposition Methods in Science and Engineering XX*, Lecture Notes in Computational Science and Engineering 91,
DOI 10.1007/978-3-642-35275-1_12,

1.1 Governing Equations

Based on [5, 16], a fundamental fuel cell model consists of five principles of conservation: mass, momentum, species, charge, and thermal energy. Typically the fuel cell is divided into seven subregions: the anode gas channel, anode gas diffusion layer (GDL), anode catalyst layer (CL), membrane, cathode gas channel, cathode GDL, and cathode CL. In the following we specifically focus our interests on mass, momentum conservation and water concentration arising in all seven subregions.

Flow equations. For flow field with velocity $\mathbf{u}$ and pressure P as unknowns, we have the following modified Navier-Stokes equations

$$\nabla \cdot (\rho \mathbf{u}) = 0, \tag{1}$$

$$\frac{1}{\varepsilon^2} \nabla \cdot (\rho \mathbf{u}\mathbf{u}) = -\nabla P + \nabla \cdot (\mu \nabla \mathbf{u}) + S_u, \tag{2}$$

where ε is porosity, ρ is density, and μ is effective viscosity. In (2) we indicate that the additional source term S_u in GDL and CL is named as Darcy's drag and defined by $S_u = -\frac{\mu}{K}\mathbf{u}$, where K is hydraulic permeability.

Species concentration equation. Water management is critical to achieve high performance for PEMFC. Therefore, without loss of generality, in order to focus on water management topics, we typically consider water as the only component in the following simplified species concentration equation. Water concentration equation in single gaseous phase is defined as follows with respect to concentration C

$$\nabla \cdot (\mathbf{u}C) = \nabla \cdot (D_g^{eff} \nabla C) + S_{H_2O}, \tag{3}$$

where $D_g^{eff} = \varepsilon^{1.5} D_{gas}$ is the effective water vapor diffusivity. The source term S_{H_2O} is given as follows.

$$S_{H_2O} = \begin{cases} -\nabla \cdot (\frac{n_d}{F}\mathbf{i}_e) - \frac{j}{2F} & \text{in cathode CL} \\ -\nabla \cdot (\frac{n_d}{F}\mathbf{i}_e) & \text{in anode CL} \\ 0 & \text{otherwise,} \end{cases} \tag{4}$$

where n_d, the electro-osmotic drag coefficient, is a constant value in our simulation. $\nabla \cdot \mathbf{i}_e = -j$ which is derived from the continuity equation of proton potential. $\mathbf{i}_e$ is the current density vector and j is the volumetric transfer current of the reaction (or transfer current density) defined by $j = j_1 - (j_1 - j_2)z/l_{cell}$. This is an approximation of transfer current density for our simplified single-phase PEMFC model due to the absence of proton and electron potentials [12].

1.2 Computational Domain and Boundary Conditions

The computational domain and its geometric sizes are schematically shown in Fig. 1 and Table 1.

For flow field (1), (2) and water concentration equation (3), the following boundary conditions are imposed:

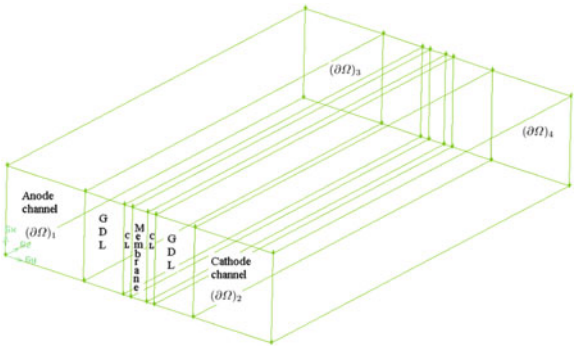


Fig. 1. Geometry of a single straight-channel PEMFC

Table 1. Physical coefficients and parameters

Parameter	Symbol	Parameter	Symbol
Anode/cathode channel width δ_{CH}	$6.180mm$	Anode/cathode GDL width δ_{GDL}	$0.235mm$
Anode/cathode CL width δ_{CL}	$0.010mm$	Membrane width δ_{mem}	$0.018mm$
Cell length l_{cell}	$70mm$	Cell depth h_{cell}	$6.360mm$
Porosity of membrane ε	0.26	Effective viscosity μ	$3.166\times10^{-5}kg/(m\cdot s)$
Porosity of GDL and CL ε	0.6	Water vapor diffusivity D_{gas}	$2.6\times10^{-5}m^2/s$
Vapor density ρ	$0.882\ kg/m^3$	Permeability of GDL and CL K	$2\times10^{-12}m^2$
Electro-osmotic drag coefficient n_d	1.5	Transfer current density j_1/j_2	$20000/10000A/m^2$

$$u_1 = u_2 = 0, u_3 = u_3|_{inlet}, C = C_{in} \quad \text{on inlet } (\partial\Omega)_1, (\partial\Omega)_2, \tag{5}$$

$$(PI - \mu\nabla\mathbf{u})\cdot\mathbf{n} = 0 \quad \text{on outlet } (\partial\Omega)_3, (\partial\Omega)_4, \tag{6}$$

$$u_1 = u_2 = u_3 = 0, \frac{\partial C}{\partial n} = 0 \quad \text{on other boundaries.} \tag{7}$$

2 Numerical Algorithm

2.1 Domain Decomposition Method and Weak Forms

First, we split the domain (Ω), shown in Fig. 1, to two overlapping subdomains: one is the anode and membrane (Ω_a), the other is the cathode and membrane (Ω_c). The interface between anode CL and membrane is denoted as $\mathscr{S}_a$, and the interface between cathode CL and membrane is denoted as $\mathscr{S}_c$. The classical overlapping Schwarz alternating method [13] is used in these two subdomains. Thus we are able to reformulate Eqs. (1)–(3) to two Dirichlet-type interfacial boundary value subproblems.

$$\text{(Problem A)}\begin{cases}\nabla\cdot(\rho\mathbf{u}_a)=0 & \text{in } \Omega_a\\ \frac{1}{\varepsilon^2}\nabla\cdot(\rho\mathbf{u}_a\mathbf{u_a})=-\nabla P_a+\nabla\cdot(\mu\nabla\mathbf{u}_a)-\frac{\mu}{K}\mathbf{u}_a & \text{in } \Omega_a\\ \nabla\cdot(\mathbf{u}_aC_a)=\nabla\cdot(D_g^{eff}\nabla C_a)+S_{H_2O} & \text{in } \Omega_a\\ u_{1,a}=u_{2,a}=0, u_{3,a}=u_3|_{inlet}, C_a=C_{a,in} & \text{on } (\partial\Omega)_1\\ (P_aI-\mu\nabla\mathbf{u_a})\cdot\mathbf{n}=0 & \text{on } (\partial\Omega)_3\\ C_a=C_c & \text{on } \mathscr{S}_c\\ u_{1,a}=u_{2,a}=u_{3,a}=0, \frac{\partial C}{\partial n}=0 & \text{on other boundaries.}\end{cases}$$

$$\text{(Problem C)}\begin{cases}\nabla\cdot(\rho\mathbf{u}_c)=0 & \text{in } \Omega_c\\ \frac{1}{\varepsilon^2}\nabla\cdot(\rho\mathbf{u}_c\mathbf{u_c})=-\nabla P_c+\nabla\cdot(\mu\nabla\mathbf{u}_c)-\frac{\mu}{K}\mathbf{u}_c & \text{in } \Omega_c\\ \nabla\cdot(\mathbf{u}_cC_c)=\nabla\cdot(D_g^{eff}\nabla C_c)+S_{H_2O} & \text{in } \Omega_c\\ u_{1,c}=u_{2,c}=0, u_{3,c}=u_3|_{inlet}, C_c=C_{c,in} & \text{on } (\partial\Omega)_2\\ (P_cI-\mu\nabla\mathbf{u_c})\cdot\mathbf{n}=0 & \text{on } (\partial\Omega)_4\\ C_c=C_a & \text{on } \mathscr{S}_a\\ u_{1,c}=u_{2,c}=u_{3,c}=0, \frac{\partial C}{\partial n}=0 & \text{on other boundaries.}\end{cases}$$

Considering various nonlinearities of equations, we particularly employ Picard's scheme to linearize the nonlinear source term. Define

$$V_a:=\{\mathbf{v}_a=(v_{1,a},v_{2,a},v_{3,a})^\top\in[H^1]^3\mid v_{1,a}|_{(\partial\Omega)_1}=v_{2,a}|_{(\partial\Omega)_1}=0, v_{3,a}|_{(\partial\Omega)_1}=u_{3,a}|_{inlet}\},$$
$$\widetilde{V}_a:=\{\mathbf{v}_a=(v_{1,a},v_{2,a},v_{3,a})^\top\in[H^1]^3\mid v_{1,a}|_{(\partial\Omega)_1}=v_{2,a}|_{(\partial\Omega)_1}=v_{3,a}|_{(\partial\Omega)_1}=0\},$$
$$Q_a:=\{w\in H^1\mid w|_{(\partial\Omega)_1}=C_{in,a}\text{ and }w|_{\mathscr{S}_c}=C_c\},\ \ \widetilde{Q}_a:=\{w\in H^1\mid w|_{(\partial\Omega)_1}=0\text{ and }w|_{\mathscr{S}_c}=0\},$$
$$P_a:=L^2(\Omega_a).$$

Then for any $(\mathbf{v}_a,q_a,w_a)\in\widetilde{V}_a\times P_a\times\widetilde{Q}_a$, find $(\mathbf{u}_a^{k+1},P_a^{k+1},C_a^{k+1})\in V_a\times P_a\times Q_a$, such that

$$\begin{cases}(\mu\nabla\mathbf{u}_a^{k+1},\nabla\mathbf{v}_a)_{\Omega_a}+(\frac{\rho}{\varepsilon^2}\nabla\mathbf{u}_a^k\mathbf{u}_a^{k+1},\mathbf{v}_a)_{\Omega_a}-(P_a^{k+1},\nabla\mathbf{v}_a)_{\Omega_a}+(\frac{\mu}{K}\mathbf{u}_a^{k+1},\mathbf{v}_a)_{\Omega_a}=0\\ (\nabla\mathbf{u}_a^{k+1},q_a)_{\Omega_a}=0\\ (D_g^{eff}\nabla C_a^{k+1},\nabla w_a)_{\Omega_a}+(\nabla\cdot(\mathbf{u}_a^kC_a),w_a)_{\Omega_a}=(S_{H_2O},w_a)_{\Omega_a},\end{cases}\tag{8}$$

which $(\cdot,\cdot)_{\Omega_i}$ stands for the L^2 inner product in Ω_i. And in subdomain Ω_c, we have the same weak form with (8).

2.2 An Overlapping Domain Decomposition Algorithm

Firstly, the subdomains Ω_a and Ω_c are partitioned into cuboids independently, which implies that the grids are local structured in anode and cathode. Define a partition $\mathscr{T}_{h_i}$ in Ω_i (i,j represent a or c), and $\Sigma_{i,j}$ is the set of mesh points of $\mathscr{T}_{h_i}$ on $\mathscr{S}_j$.

To discretize weak form (8), we introduce the finite element space $V_{h_i}\times P_{h_i}\subseteq V_i\times P_i$ on $\mathscr{T}_{h_i}$, where $V_{h_i}\times P_{h_i}$ denotes the $Q2Q1$ (triquadratic velocity and trilinear pressure) finite element spaces. Q_{h_a} denotes the triquadratic finite element space for water concentration whose members equal f_a on $\mathscr{S}_c$, where f_a represents the values of points in the sets of $\Sigma_{a,c}$, which are obtained from the previous alternating step C^k

by lagrange interpolation. Moreover, let $\widetilde{Q}_{h_a} \subseteq \widetilde{Q}_a$ be the triquadratic finite element space and $\widetilde{V}_{h_a} \subseteq \widetilde{V}_a$ be the triquadratic finite element space. In subdomain Ω_c, Q_{h_c} and $\widetilde{V}_{h_c}$ are defined in the same ways.

For flow and water concentration equations, we introduce the following combined finite element-upwind finite volume schemes [11].

For any given $(\mathbf{u}_{h_i}^k, P_{h_i}^k, C_{h_i}^k) \in V_{h_i} \times P_{h_i} \times Q_{h_i}$ $(k = 0, 1, 2, \ldots)$, find $(\mathbf{u}_{h_i}^{k+1}, P_{h_i}^{k+1}, C_{h_i}^{k+1}) \in V_{h_i} \times P_{h_i} \times Q_{h_i} (k = 0, 1, 2, \ldots)$, such that

$$(\mu \nabla \mathbf{u}_{h_i}^{k+1}, \nabla \mathbf{v}_{h_i})_{\Omega_i} + (\frac{\rho}{\varepsilon^2} \nabla \mathbf{u}_{h_i}^k \mathbf{u}_{h_i}^{k+1}, \mathbf{v}_{h_i})_{\Omega_i} - (P_{h_i}^{k+1}, \nabla \mathbf{v}_{h_i})_{\Omega_i} + (\frac{\mu}{K} \mathbf{u}_{h_i}^{k+1}, \mathbf{v}_{h_i})_{\Omega_i} = 0$$
$$(\nabla \mathbf{u}_{h_i}^{k+1}, q_{h_i})_{\Omega_i} = 0 \quad \forall (\mathbf{v}_{h_i}, q_{h_i}) \in \widetilde{V}_{h_i} \times P_{h_i}, \quad (9)$$
$$(D_g^{eff} \nabla C_{h_i}^{k+1}, \nabla w_{h_i})_{\Omega_i} + (\nabla \cdot (\mathbf{u}_{h_i}^{k+1} C_{h_i}^{k+1}), w_{h_i})_{\Omega_i} + \delta(h_i) \mathbf{u}_{h_i}^{k+1} \cdot (\nabla C_{h_i}^{k+1}, \nabla w_{h_i})_{\Omega_i}$$
$$= (S_{H_2O}, w_{h_i})_{\Omega_i} \quad \forall w_{h_i} \in \widetilde{Q}_{h_i}, \quad (10)$$

where the last term in the left hand side of (10) is a stabilizing term, derived from streamline-diffusion scheme [3, 6–8]. Basically we hold $\delta(h) = Ch$, C is a certain constant parameter, which is chosen artificially with least possible on the premise of optimal stability. Usually starting with small ones, we gradually increase the value of C and compute the corresponding finite element equation (10) until gained numerical solutions are not oscillating any more in convection-dominated gas channel.

Now, we are in position to describe the overlapping domain decomposition algorithm with the finite element discretizations.

Algorithm: Given $\mathbf{u}_h^0, C_h^0$, the following procedures are successively executed $(k > 0)$:

Step 1. Solve (9) in Ω_a and Ω_c for $(\mathbf{u}_{h_i}^{k+1}, P_{h_i}^{k+1})$, respectively, until

$$\|\mathbf{u}_{h_i}^{k+1} - \mathbf{u}_{h_i}^k\|_{L^2(\Omega_i)} + \|P_{h_i}^{k+1} - P_{h_i}^k\|_{L^2(\Omega_i)} < \text{tolerance}. \quad (11)$$

Step 2. Solve (10) for $C_{h_a}^{k+1}$, and construct the finite element space $\widetilde{Q}_{h_c}$ for Ω_c.

Step 3. Solve (10) for $C_{h_c}^{k+1}$, and construct the finite element space $\widetilde{Q}_{h_a}$ for Ω_a.

Step 4. Compute the following stopping criteria:

$$\|C_{h_a}^{k+1} - C_{h_a}^k\|_{L^2(\Omega_a)} < \text{tolerance}. \quad (12)$$

If yes, then numerical computation is complete. Otherwise, go back to the step 2 and continue.

3 Numerical Results

In this section, we will carry out the following numerical experiments which indicate that our methods are effective to deal with the non-matching grids, see Fig. 2 for example, in the simulation of the PEMFC. The velocity $u_3|_{inlet}$ is defined as a paraboloidal-like function given in (13).

$$u_3|_{inlet} = \begin{cases} 0.2\sin\frac{x\pi}{\delta_{CH}}\sin\frac{y\pi}{\delta_{CH}} & \text{on anode inlet } (\partial\Omega)_1 \\ 0.3\sin\frac{x\pi}{\delta_{CH}}\sin\frac{(y-l_{add})\pi}{\delta_{CH}} & \text{on cathode inlet } (\partial\Omega)_2 \end{cases}, \tag{13}$$

where $l_{add} = \delta_{CH} + \delta_{GDL} + \delta_{CL} + \delta_{mem}$.

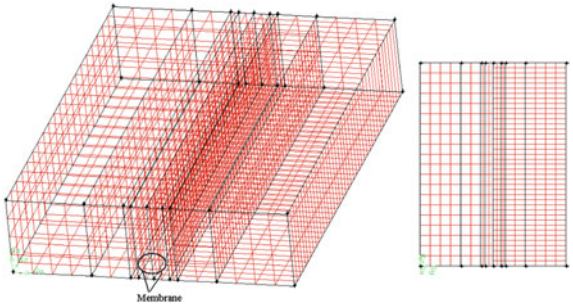


Fig. 2. An example of non-matching grids

Figures 3 and 4 show the velocity field in anode and cathode of fuel cell at the face of $x = 3.18$ mm with this two method. As expected, there is a large difference in the velocity scale between the porous media and the open channel. The velocity in porous GDL is at least two orders of magnitude smaller than that in the open gas channel, indicating that gas diffusion is the dominant transport mechanism in porous GDL. Porous CL has a smaller velocity than GDL due to the inferior diffusion ability.

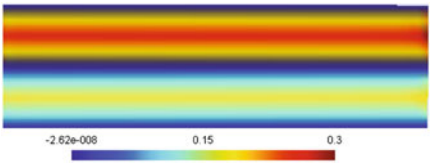


Fig. 3. Velocity with DDM

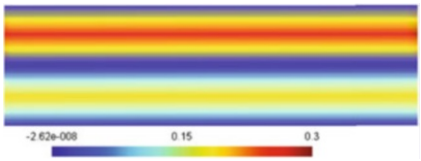


Fig. 4. Velocity with single domain

Figure 5 displays the water concentration distribution, presenting in the phase of water vapor, in anode and cathode. As shown in the figure, significant variations are displayed in both anode and cathode; in the porous media there is an increased water vapor concentration along the channel.

In order to verify the correctness of our numerical solutions, we compute the relative error of mass balance in terms of the numerical fluxes at the inlet and outlet.

$$\text{mass balance error} = \frac{|\int_{(\partial\Omega)_{outlet}} Cu_3 dS - \int_{(\partial\Omega)_{inlet}} C_{in}u_3|_{inlet} dS - \int_{\Omega} S_{H_2O} dV|}{\int_{(\partial\Omega)_{inlet}} C_{in}u_3|_{inlet} dS.} \tag{14}$$

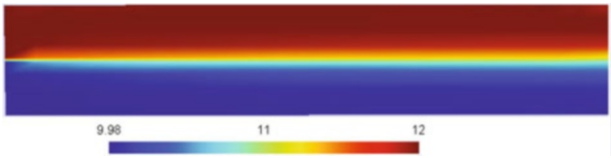


Fig. 5. Distributions of water concentration with DDM

The tolerance of our stopping criteria (12) for Schwarz alternating iteration is 10^{-20}. By plugging the assigned and the computed concentration C as well as horizontal velocity u_3 in Eq. (14), we attain a convergent mass balance error for our numerical solutions along with the continuously refining grids, shown in Table 2. A more accurate mass balance error is attained for the numerical solutions with DDM.

Table 2. Convergent mass balance error for with different grids

	Grids	Unknowns	Error with DDM	Error with single domain
Mesh1	720	36260	9.731×10^{-3}	8.112×10^{-3}
Mesh2	1440	58660	8.338×10^{-3}	6.909×10^{-3}
Mesh3	2880	115884	3.774×10^{-3}	2.233×10^{-3}
Mesh4	3600	139840	1.528×10^{-3}	Overflow

4 Conclusions and Future Work

In this paper, a simplified single-phase 3D steady PEMFC model is introduced by a modified Navier-Stokes equations for mass and momentum, and a conservation equation for water concentration. Based on the combined finite element-upwind finite volume methods and the overlapping domain decomposition method, a new discretization scheme is designed and implemented for the PEMFC model. Numerical experiments demonstrate that our methods are effective to deal with the non-matching grids and obtain a relatively accurate numerical solution with low mass balance error. The derived discretization scheme will be also studied for two-phase unsteady and/or fuel cell stack model in our further work.

Acknowledgments The support from the NFSC (No.11101311) is fully acknowledged. Ziping Huang also acknowledge "Applied Mathematics Chair Fund of China-German College" (0900101021). Pengtao Sun is supported by NSF Grant DMS-0913757.

Bibliography

[1] R. E. Bank and J. Xu. Asymptotically exact a posteriori error estimators, part I: grids with superconvergence. *SIAM Journal on Numerical Analysis*, 41:2294–2312, 2003.

[2] C. H Cheng and H. H Linb. Numerical analysis of effects of flow channel size on reactant transport in a proton exchange membrane fuel cell stack. *Journal of Power Sources*, 194:349–359, 2009.

[3] M. Feistauer and J. Felcman. On the convergence of a combined finite volume-finite element for nonlinear convection–diffusion problems. *Numerical Methods for Partial Differential Equations*, 13:163–190, 1997.

[4] Y. Q. Huang. Superconvergence for quadratic triangular finite elements on mildly structured grids. *Mathematics of computation*, 77:1253–1268, 2008.

[5] Hyunchul Ju A single-phase, non-isothermal model for pem fuel cells. *International Journal of Heat and Mass Transfer*, 48:1303–1315, 2005.

[6] Dietmar Kroner and Mirko Ohlbefger. A posteriori error estimates for upwind finite volume schemes for nonlinear conservation laws in multi-dimensions. *Numerische Mathematik*, 69:25–39, 2000.

[7] Dietmar Kroner and Mirko Rokyta. Convergence of upwind finite volume schemes for scalar conservation laws in two dimensions. *SIAM Journal on Numerical Analysis*, 31:324–343, 1994.

[8] Dietmar Kroner. Convergence of higher order upwind finite volume schemes on unstructured grids for scalar conservation laws in several space dimensions. *Mathematics of Computation*, 71:527–560, 1995.

[9] Atul Kumar and Ramana G. Reddy. Effect of channel dimensions and shape in the flow-field distributor on the performance of polymer electrolyte membrane fuel cells. *Journal of Power Sources*, 113:11–18, 2003.

[10] S. Shimpalee and J.W. Van Zee. Numerical studies on rib and channel dimension of flow-field on PEMFC performance. *International Journal of Hydrogen Energy*, 32:842–856, 2007.

[11] Pengtao Sun, Guangri Xue, Chaoyang Wang, and Jinchao Xu. A domain decomposition method for two-phase transport model in the cathode of a polymer electrolyte fuel cel. *Journal of Computational Physics*, 228:6016–6036, 2009.

[12] Pengtao Sun, Guangri Xue, Chaoyang Wang, and Jinchao Xu. Fast numerical simulation of two-phase transport model in the cathode of a polymer electrolyte fuel cell. *Communications in Computational Physics*, 6:49–71, 2009.

[13] Andrea Toselli and Olof B.Widlund. *Domain decomposition methods - algorithms and theory*. Springer, New York, 2005.

[14] L. B. Wahlbin. *Superconvergence in Galerkin Finite Element Methods*. Springer, Berlin, 1995.

[15] C. Y. Wang. Computational fluid dynamics modeling of proton exchange membrane fuel cells. *Journal of the Electrochemical Society*, 147:4485–4493, 2000.

[16] C. Y. Wang. Fundamental models for fuel cell engineering. *Journal of the Electrochemical Society*, 104:4727–4766, 2004.

Multigrid Methods for the Biharmonic Problem with Cahn-Hilliard Boundary Conditions

Susanne C. Brenner[1], Shiyuan Gu[2], and Li-yeng Sung[3]

[1] Department of Mathematics and Center for Computation & Technology, Louisiana State University, Baton Rouge, LA 70803, USA. brenner@math.lsu.edu
[2] Department of Mathematics and Center for Computation & Technology, Louisiana State University, Baton Rouge, LA 70803, USA. gshy@math.lsu.edu
[3] Department of Mathematics and Center for Computation & Technology, Louisiana State University, Baton Rouge, LA 70803, USA. sung@math.lsu.edu

1 Introduction

Let $\Omega \subset \mathbb{R}^2$ be a bounded polygonal domain, $V = \{v \in H^2(\Omega) : \partial v/\partial n = 0 \text{ on } \partial\Omega\}$ and $f \in L_2(\Omega)$. In this paper we consider multigrid methods for the following biharmonic problem: Find $u \in V$ such that

$$\int_\Omega \nabla^2 u : \nabla^2 v \, dx = \int_\Omega f v \, dx \qquad \forall v \in V, \tag{1}$$

where $\nabla^2 w : \nabla^2 v = \sum_{i,j=1}^2 w_{x_i x_j} v_{x_i x_j}$ is the inner product of the Hessian matrices of w and v. Under the (assumed) compatibility condition,

$$\int_\Omega f \, dx = 0, \tag{2}$$

the biharmonic problem (1) is solvable and the solution is unique up to an additive constant. Furthermore we have an elliptic regularity estimate

$$\|\hat{u}\|_{H^{2+\alpha}(\Omega)} \le C \|f\|_{L_2(\Omega)} \tag{3}$$

for the solution $\hat{u}$ of (1) that satisfies $\int_\Omega \hat{u} \, dx = 0$. Note that, unlike the biharmonic problem with the boundary conditions of clamped plates, the index of elliptic regularity α in (3), which is determined by the angles of Ω, can be close to 0 even if Ω is convex (cf. [2]).

The essential boundary condition $\partial u/\partial n = 0$ and the natural boundary condition $\partial(\Delta u)/\partial n = 0$ satisfied by the solution u of (1) appear in the Cahn-Hilliard model for phase separation phenomena (cf. [8]). In particular, the boundary value problem (1) appears when the Cahn-Hilliard equation is discretized in time by an implicit method and the resulting nonlinear fourth order elliptic boundary value problem is solved by an Newton iteration.

R. Bank et al. (eds.), *Domain Decomposition Methods in Science and Engineering XX*, Lecture Notes in Computational Science and Engineering 91,
DOI 10.1007/978-3-642-35275-1_13,

We will describe a C^0 interior penalty method for (1) in Sect. 2 and introduce in Sect. 3 multigrid methods that are based on a new smoother. The convergence properties of the multigrid methods are briefly discussed in Sect. 4, followed by numerical results in Sect. 5.

2 A Quadratic C^0 Interior Penalty Method

C^0 interior penalty methods (cf. [6, 9]) are discontinuous Galerkin methods for fourth order problems. Let $\mathscr{T}_h$ be a simplicial triangulation of Ω, $V_h \subset H^1(\Omega)$ be the associated P_2 Lagrange finite element space (cf. [5]), and $\hat{V}_h$ be the subspace of V_h consisting of functions with zero mean, i.e., $v \in V_h$ belongs to $\hat{V}_h$ if and only if $\int_\Omega v\,dx = 0$. The quadratic C^0 interior penalty method for (1) is to find $\hat{u}_h \in \hat{V}_h$ such that

$$a_h(\hat{u}_h, v) = \int_\Omega f v\,dx \qquad \forall\, v \in \hat{V}_h, \tag{4}$$

where

$$\begin{aligned} a_h(w,v) = &\sum_{T\in\mathscr{T}_h}\int_T \nabla^2 w : \nabla^2 v\,dx + \sum_{e\in\mathscr{E}_h}\int_e \left\{\!\!\left\{ \frac{\partial^2 w}{\partial n^2} \right\}\!\!\right\} \left[\!\!\left[\frac{\partial v}{\partial n} \right]\!\!\right] ds \\ &+ \sum_{e\in\mathscr{E}_h}\int_e \left\{\!\!\left\{ \frac{\partial^2 v}{\partial n^2} \right\}\!\!\right\} \left[\!\!\left[\frac{\partial w}{\partial n} \right]\!\!\right] ds + \sum_{e\in\mathscr{E}_h} \frac{\sigma}{|e|}\int_e \left[\!\!\left[\frac{\partial w}{\partial n} \right]\!\!\right] \left[\!\!\left[\frac{\partial v}{\partial n} \right]\!\!\right] ds. \end{aligned} \tag{5}$$

Here $\mathscr{E}_h$ is the set of the edges in $\mathscr{T}_h$, $\{\!\{\partial^2 v/\partial n^2\}\!\}$ (resp. $[\![\partial v/\partial n]\!]$) is the average of the second normal derivative of v (resp. the jump of the first normal derivative of v) across an edge, $|e|$ is the length of the edge e, and $\sigma > 0$ is a penalty parameter.

The quadratic C^0 interior penalty method is consistent. It is also stable if σ is sufficiently large, which is assumed to be the case. (The magnitude of σ is related to certain inverse estimates. It can be taken to be 5 in practice.) It can be shown (cf. [3]) that the solution $\hat{u}_h$ of (4) satisfies the following error estimate:

$$\|\hat{u} - \hat{u}_h\|_h \le Ch^\alpha \|f\|_{L_2(\Omega)}, \tag{6}$$

where $\hat{u}$ is the zero mean solution of (1), α is the index of elliptic regularity in (3), and the norm $\|\cdot\|_h$ is given by

$$\|v\|_h^2 = \sum_{T\in\mathscr{T}_h} |v|_{H^2(T)}^2 + \sum_{e\in\mathscr{E}_h} |e|^{-1} \|[\![\partial v/\partial n]\!]\|_{L_2(e)}^2.$$

C^0 interior penalty methods have certain advantages over other finite element methods for fourth order problems. They are simpler than conforming methods which require C^1 elements. They come in a natural hierarchy that can capture smooth solutions efficiently, which is not the case for classical nonconforming methods. Unlike mixed methods they preserve the positive definiteness of the continuous problem and are easier to develop for more complicated problems (cf. [9]).

Another significant advantage of C^0 interior penalty methods comes from the fact that the underlying finite element spaces are standard spaces for second order problems. (Note that the essential boundary condition for (1) is only enforced weakly in (4) and the finite element space V_h does not involve any boundary condition.) Therefore multigrid solves for second order problems can be easily implemented as a preconditioner. By using such a preconditioner in the smoothing steps of multigrid algorithms for fourth order problems, the performance of the smoother and hence the overall performance of the multigrid algorithms can be significantly improved. This approach was carried out in [7] for the biharmonic problem with the boundary conditions of clamped plates. Below we will use this approach to develop multigrid methods for (4).

3 Multigrid Methods

Let $\mathscr{T}_k$ $(k = 0, 1, \cdots)$ be a sequence of simplicial triangulations obtained from the initial triangulation $\mathscr{T}_0$ by uniform refinement. We will use V_k (resp. $a_k(\cdot,\cdot)$) to denote the finite element space (resp. the bilinear form for the C^0 interior penalty method) associated with $\mathscr{T}_k$.

Let V_k' be the dual space of V_k and $\hat{V}_k = \{v \in V_k : \int_\Omega v\,dx = 0\}$ be the zero-mean subspace of V_k. We can identify $\hat{V}_k'$ with the subspace of V_k' whose members annihilate the constant functions, i.e., $\hat{V}_k' = \{\gamma \in V_k' : \langle \gamma, 1\rangle = 0\}$, where $\langle\cdot,\cdot\rangle$ is the canonical bilinear form between a vector space and its dual.

Let the operator $A_k : V_k \longrightarrow \hat{V}_k'$ be defined by $\langle A_k v, w\rangle = a_k(v, w)$ for all $v, w \in V_k$. We can then rewrite the discrete problem (4) as $A_k \hat{u}_k = \phi_k$, where $\hat{u}_k \in \hat{V}_k$ and $\phi_k \in \hat{V}_k'$ satisfies $\langle \phi_k, v\rangle = \int_\Omega f v\,dx$ for all $v \in V_k$. Below we will develop multigrid algorithms for equations of the form

$$A_k z = \psi \tag{7}$$

where $z \in \hat{V}_k$ and $\psi \in \hat{V}_k'$.

There are two ingredients in the design of multigrid algorithms. First of all, we need intergrid transfer operators to move data between consecutive levels. Since the finite element spaces are nested, we can take the coarse-to-fine operator $I_{k-1}^k : V_{k-1} \longrightarrow V_k$ to be the natural injection and the fine-to-coarse operator $I_k^{k-1} : V_k' \longrightarrow V_{k-1}'$ to be the transpose of I_{k-1}^k with respect to the canonical bilinear forms, i.e., $\langle I_k^{k-1}\gamma, v\rangle = \langle \gamma, I_{k-1}^k v\rangle$ for all $\gamma \in V_k'$, $v \in V_{k-1}$. Note that I_{k-1}^k maps $\hat{V}_{k-1}$ into $\hat{V}_k$ and consequently I_k^{k-1} maps $\hat{V}_k'$ into $\hat{V}_{k-1}'$.

The second ingredient is a good smoother that can damp out the highly oscillatory part of the error of an approximate solution so that the remaining part of the error can be captured accurately on a coarser grid. Here we take advantage of the fact that the P_2 Lagrange finite element space is a standard space for second order problems to incorporate a multigrid Poisson solve in the smoother. Let $L_k : \hat{V}_k \longrightarrow \hat{V}_k'$ be the discrete Laplace operator defined by

$$\langle L_k v, w\rangle = \int_\Omega \nabla v \cdot \nabla w\,dx \qquad \forall v, w \in \hat{V}_k.$$

We take $S_k^{-1}: \hat{V}_k' \longrightarrow \hat{V}_k$ to be an approximate inverse of L_k obtained from a multigrid Poisson solve such that

$$\langle S_k v, v\rangle \approx |v|^2_{H^1(\Omega)} \qquad \forall\, v \in \hat{V}_k. \tag{8}$$

The smoothing step in our multigrid algorithms for (7) is then given by

$$z_{\text{new}} = z_{\text{old}} + \lambda_k S_k^{-1}(\psi - A_k z_{\text{old}}), \tag{9}$$

where λ_k is a damping factor chosen so that the spectral radius $\rho(\lambda_k S_k^{-1} A_k)$ is <2. It follows from (8) and standard inverse estimates (cf. [5]) that we can take $\lambda_k = Ch_k^2$. Note that the computational cost of (9) is proportional to the dimension of $\hat{V}_k$, which implies that the overall computational costs of the multigrid algorithms in Sects. 3.1 and 3.2 are also proportional to the dimension of $\hat{V}_k$.

We can now describe the V-cycle and W-cycle algorithms (cf. [10, 11]) in terms of the intergrid transfer operators and the smoothing scheme.

3.1 V-Cycle Algorithm

The V-cycle algorithm computes an approximate solution $MG_V(k,\psi,z_0,m)$ of (7) with initial guess $z_0 \in \hat{V}_k$ and m pre-smoothing and m post-smoothing steps. For $k=0$, we take $MG_V(0,\psi,z_0,m)$ to be the output of a direct solve. For $k \geq 1$, we compute $MG_V(k,\psi,z_0,m)$ recursively in three steps.

Pre-smoothing For $1 \leq \ell \leq m$, compute z_ℓ recursively by

$$z_\ell = z_{\ell-1} + \lambda_k S_k^{-1}(\psi - A_k z_{\ell-1}).$$

Coarse Grid Correction Compute

$$z_{m+1} = z_m + I_{k-1}^k MG_V(k-1,\rho_{k-1},0,m),$$

where $\rho_{k-1} = I_k^{k-1}(\psi - A_k z_m) \in \hat{V}_{k-1}'$ is the transferred residual of z_m.

Post-smoothing For $m+2 \leq \ell \leq 2m+1$, compute z_ℓ recursively by

$$z_\ell = z_{\ell-1} + \lambda_k S_k^{-1}(\psi - A_k z_{\ell-1}).$$

The final output is $MG_V(k,\psi,z_0,m) = z_{2m+1}$.

3.2 W-Cycle Algorithm

The W-cycle algorithm computes an approximate solution $MG_W(k,\psi,z_0,m)$ of (7) with initial guess $z_0 \in \hat{V}_k$ and m pre-smoothing and m post-smoothing steps. The only difference between the V-cycle algorithm and the W-cycle algorithm is in the coarse grid correction step, where the coarse grid algorithm is applied twice to the coarse grid residual equation. More precisely, we have

$$\begin{aligned} z_{m+\frac{1}{2}} &= MG_W(k-1,\rho_{k-1},0,m),\\ z_{m+1} &= z_m + MG_W(k-1,\rho_{k-1},z_{m+\frac{1}{2}},m). \end{aligned}$$

Remark 1. For simplicity we have described the multigrid algorithms in terms of the space $\hat{V}_k$ where the bilinear form $a_k(\cdot,\cdot)$ is nonsingular. But the multigrid Poisson solve S_k^{-1} (and hence the V-cycle and W-cycle algorithms) can be implemented on V_k for $k \geq 1$. The implementation of multigrid algorithms for the singular Neumann problem is discussed for example in [1].

4 Convergence Properties

Let $z_0 \in \hat{V}_k$ be the initial guess and $z_\dagger \in \hat{V}_k$ be the output of the V-cycle or W-cycle algorithm for (7). Numerical results indicate that

$$\|z - z_\dagger\|_{a_h} \leq Cm^{-\alpha}\|z - z_0\|_{a_h}, \tag{10}$$

where α is the index of elliptic regularity in (3) and $\|\cdot\|_{a_h} = \sqrt{a_h(\cdot,\cdot)}$ is the energy norm, provided that the number of smoothing steps $m \geq m_*$. Here m_* is a sufficiently large positive integer independent of k. In particular the multigrid algorithms are contractions for sufficiently large m and the contraction numbers are bounded away from 1 uniformly. A similar estimate was obtained in [7] for the boundary conditions of clamped plates. The derivation of (10) for the Cahn-Hilliard boundary conditions will be carried out in [4] where general fourth order problems are considered.

A significant benefit of including a multigrid Poisson solve in the smoothing step (9) is that the resulting smoothing property is similar to that for second order problems (cf. [7]) so that the contraction number estimate (10) is also similar to that for second order problems. Indeed, because of the estimate (8), we can derive a smoothing property for (9) with respect to a family of mesh dependent norms $|||\cdot|||_{s,k}$ such that $|||\cdot|||_{0,k} \approx |\cdot|_{H^1(\Omega)}$ and $|||\cdot|||_{1,k} \approx |\cdot|_{H^2(\Omega)}$ on the space $\hat{V}_k$. Note that the smoothing properties of standard smoothers for second order problems are described in terms of mesh dependent norms $|||\cdot|||_{s,k}$ such that $|||\cdot|||_{0,k} \approx \|\cdot\|_{L_2(\Omega)}$ and $|||\cdot|||_{1,k} \approx |\cdot|_{H^1(\Omega)}$ on the finite element spaces. The good performance of the smoothing step (9) is due to the similarity between the Hilbert scales $[H^1(\Omega), H^2(\Omega)]$ and $[L_2(\Omega), H^1(\Omega)]$.

If we use a standard smoother such as the Richardson relaxation in a multigrid algorithm for (7), then the smoothing property will be determined by the Hilbert scale $[L_2(\Omega), H^2(\Omega)]$. In this case the estimate (10) will be replaced by the estimate

$$\|z - z_\dagger\|_{a_h} \leq Cm^{-\alpha/2}\|z - z_0\|_{a_h}, \tag{11}$$

which means that the effect of 100 smoothing steps without the preconditioner is roughly equivalent to the effect of 10 smoothing steps with the preconditioner. As far as we know, all existing multigrid methods for fourth order problems (except those in [6]) use standard smoothers and their convergence is governed by (11).

5 Numerical Results

The numerical experiments were performed on sienna@IMA (Intel P4, 3.4 GHz CPU, 2 G memory) at the Institute for Mathematics and its Applications. In the numerical experiments we take $\sigma = 5$ and the preconditioner to be a V-cycle Poisson

solve with one pre-smoothing step and one post-smoothing step. (Other multigrid Poisson solves can also be used, but the $V(1,1)$ solve appears to be the most efficient.) The contraction numbers for the V-cycle and W-cycle algorithms on the unit square (with two elements in the initial mesh) are reported in Tables 1 and 2. It is observed that the V-cycle (resp. W-cycle) algorithm is a contraction for $m \geq 4$ (resp. $m \geq 2$).

Table 1. Contraction numbers for the V-cycle algorithm on the unit square.

k \ m	4	5	6	7	8	9	10	11	12	13
1	0.212	0.126	0.0813	0.0594	0.0442	0.0332	0.0252	0.0192	0.0147	0.0114
2	0.329	0.223	0.190	0.164	0.142	0.124	0.109	0.0967	0.0861	0.0771
3	0.412	0.342	0.308	0.279	0.255	0.234	0.217	0.203	0.190	0.179
4	0.479	0.420	0.386	0.357	0.334	0.314	0.296	0.282	0.266	0.257
5	0.537	0.467	0.434	0.408	0.386	0.367	0.351	0.336	0.324	0.312
6	0.578	0.494	0.462	0.436	0.415	0.396	0.380	0.366	0.353	0.341
7	0.619	0.503	0.472	0.446	0.425	0.406	0.391	0.376	0.364	0.351

Table 2. Contraction numbers for the W-cycle algorithm on the unit square.

k \ m	2	3	4	5	6	7	8	9	10	11
1	0.661	0.368	0.212	0.126	0.0813	0.0594	0.0442	0.0332	0.0252	0.0192
2	0.483	0.360	0.291	0.241	0.203	0.172	0.148	0.128	0.112	0.0983
3	0.475	0.375	0.335	0.282	0.263	0.229	0.215	0.195	0.182	0.171
4	0.455	0.383	0.335	0.308	0.287	0.270	0.256	0.244	0.233	0.223
5	0.456	0.384	0.344	0.315	0.297	0.279	0.267	0.255	0.245	0.237
6	0.455	0.384	0.344	0.316	0.297	0.280	0.268	0.256	0.248	0.239
7	0.455	0.384	0.344	0.317	0.297	0.281	0.269	0.258	0.248	0.240

For comparison we report in Table 3 the contraction numbers for the V-cycle algorithm that does not use a preconditioner in the smoothing steps. The smoothing step in this algorithm is the standard Richardson relaxation scheme.

We have also carried out numerical experiments for the L-shaped domain with vertices $(0,0)$, $(1,0)$, $(1,1)$, $(-1,1)$, $(-1,-1)$ and $(0,-1)$. The initial mesh consists of six isosceles triangles sharing $(0,0)$ as a common vertex. The contraction numbers for the W-cycle algorithm with/without the preconditioner are presented in Tables 4 and 5.

We note that the contraction numbers in Table 1 (resp. Table 4) for m smoothing steps are comparable to the contraction numbers in Table 3 (resp. Tables 5) for m^2 smoothing steps.

Table 3. Contraction numbers for the V-cycle algorithm without a preconditioner on the unit square.

k \ m	21	22	23	24	25	26	27	28	29	30
1	0.428	0.410	0.392	0.376	0.361	0.346	0.332	0.320	0.307	0.296
2	0.646	0.614	0.583	0.555	0.529	0.504	0.481	0.459	0.439	0.420
3	0.770	0.728	0.690	0.654	0.621	0.591	0.562	0.535	0.510	0.487
4	0.844	0.797	0.753	0.713	0.676	0.641	0.609	0.579	0.551	0.525
5	0.895	0.843	0.795	0.752	0.711	0.674	0.639	0.607	0.577	0.548
6	0.931	0.876	0.826	0.780	0.737	0.697	0.661	0.627	0.595	0.565
7	0.960	0.902	0.849	0.801	0.757	0.715	0.677	0.642	0.609	0.578

Table 4. Contraction numbers for the W-cycle algorithm with a preconditioner on the L-shaped domain.

k \ m	3	5	7	9	11	13	15	17	19	21	23
1	0.319	0.187	0.125	0.105	0.0913	0.0798	0.0699	0.0614	0.0540	0.0476	0.0420
2	0.383	0.273	0.206	0.161	0.139	0.132	0.125	0.119	0.113	0.108	0.103
3	0.390	0.302	0.238	0.208	0.182	0.163	0.152	0.148	0.144	0.141	0.137
4	0.386	0.309	0.271	0.245	0.224	0.208	0.193	0.181	0.170	0.161	0.153
5	0.384	0.315	0.279	0.255	0.237	0.222	0.209	0.198	0.189	0.180	0.172
6	0.384	0.316	0.281	0.257	0.240	0.226	0.213	0.203	0.193	0.185	0.177
7	0.387	0.317	0.281	0.258	0.240	0.226	0.214	0.203	0.194	0.186	0.178

Table 5. Contraction numbers for the W-cycle algorithm without a preconditioner on the L-shaped domain.

k \ m	5	7	9	11	13	15	17	19	21	23
1	0.943	0.788	0.680	0.600	0.537	0.486	0.443	0.407	0.375	0.347
2	0.790	0.585	0.505	0.459	0.426	0.394	0.375	0.358	0.342	0.328
3	0.666	0.512	0.469	0.456	0.434	0.416	0.400	0.386	0.373	0.362
4	0.580	0.519	0.484	0.454	0.434	0.418	0.405	0.394	0.385	0.376
5	0.581	0.527	0.491	0.465	0.444	0.427	0.414	0.402	0.392	0.384
6	0.587	0.531	0.494	0.467	0.446	0.429	0.415	0.404	0.394	0.386
7	0.587	0.530	0.493	0.467	0.446	0.429	0.415	0.404	0.394	0.386

Finally we compare the computational cost between the preconditioned schemes and the un-preconditioned schemes. On the unit square, the contraction numbers for the preconditioned V-cycle algorithm with $m = 4$ (cf. Table 1) are about the same as the contraction numbers for the un-preconditioned V-cycle algorithm with $m = 29$ (cf. Table 3). For $k = 7$, the former takes 1.4×10^8 floating point operations and 0.55 s while the latter takes 3.2×10^8 floating point operations and 1.2 s.

On the L-shaped domain, the contraction numbers for the preconditioned W-cycle algorithm with $m = 3$ (cf. Table 4) are about the same as the contraction numbers for the un-preconditioned W-cycle algorithm with $m = 23$ (cf. Table 5). For $k = 7$, the former takes 4.7×10^8 floating point operations and 2.1 s while the latter takes 1.1×10^9 floating point operations and 4.7 s.

Acknowledgments This work was supported in part by the National Science Foundation under Grant No. DMS-10-16332 and by the Institute for Mathematics and its Applications with funds provided by the National Science Foundation.

Bibliography

[1] R.E. Bank and T.F. Dupont. An optimal order process for solving finite element equations. *Math. Comp.*, 36:35–51, 1981.

[2] H. Blum and R. Rannacher. On the boundary value problem of the biharmonic operator on domains with angular corners. *Math. Methods Appl. Sci.*, 2:556–581, 1980.

[3] S.C. Brenner, S. Gu, T. Gudi, and L.-Y. Sung. A C^0 interior penalty method for a biharmonic problem with essential and natural boundary conditions of Cahn-Hilliard type. *preprint*, 2010.

[4] S.C. Brenner, S. Gu, and L.-Y. Sung. Multigrid methods for fourth order problems with essential and natural boundary conditions of the Cahnh-Hilliard type. (*in preparation*)

[5] S.C. Brenner and L.R. Scott. *The Mathematical Theory of Finite Element Methods* (*Third Edition*). Springer-Verlag, New York, 2008.

[6] S.C. Brenner and L.-Y. Sung. C^0 interior penalty methods for fourth order elliptic boundary value problems on polygonal domains. *J. Sci. Comput.*, 22/23: 83–118, 2005.

[7] S.C. Brenner and L.-Y. Sung. Multigrid algorithms for C^0 interior penalty methods. *SIAM J. Numer. Anal.*, 44:199–223, 2006.

[8] J.W. Cahn and J.E. Hilliard. Free energy of a nonuniform system-I: Interfacial free energy. *J. Chem. Phys.*, 28:258–267, 1958.

[9] G. Engel, K. Garikipati, T.J.R. Hughes, M.G. Larson, L. Mazzei, and R.L. Taylor. Continuous/discontinuous finite element approximations of fourth order elliptic problems in structural and continuum mechanics with applications to thin beams and plates, and strain gradient elasticity. *Comput. Methods Appl. Mech. Engrg.*, 191:3669–3750, 2002.

[10] W. Hackbusch. *Multi-grid Methods and Applications*. Springer-Verlag, Berlin-Heidelberg-New York-Tokyo, 1985.

[11] J. Mandel, S. McCormick, and R. Bank. Variational Multigrid Theory. In S. McCormick, editor, *Multigrid Methods, Frontiers In Applied Mathematics 3*, pages 131–177. SIAM, Philadelphia, 1987.

A Two-Level Additive Schwarz Preconditioner for C^0 Interior Penalty Methods for Cahn-Hilliard Equations

Kening Wang

University of North Florida, 1 UNF Drive, Jacksonville, FL 32224
kening.wang@unf.edu

Summary. We study a two-level additive Schwarz preconditioner for C^0 interior penalty methods for a biharmonic problem with essential and natural boundary conditions with Cahn-Hilliard type. We show that the condition number of the preconditioned system is bounded by $C(1+(H^3/\delta^3))$, where H is the typical diameter of a subdomain, δ measures the overlap among the subdomains, and the positive constant C is independent of the mesh sizes and the number of subdomains.

1 Introduction

Let Ω be a bounded polygonal domain in $\mathbb{R}^2$, and $\mathbb{V} = \{v \in H^2(\Omega) : \partial v/\partial n = 0 \text{ on } \partial\Omega\}$, where $\partial/\partial n$ denotes the outward normal derivative. Consider the following model problem which is the weak form of the biharmonic problem with boundary conditions of Cahn-Hilliard type:

Find $u \in H^2(\Omega)$ such that

$$a(u,v) = (f,v) \qquad \forall v \in \mathbb{V}, \tag{1}$$

$$\frac{\partial u}{\partial n} = 0 \qquad \text{on } \partial\Omega, \tag{2}$$

where $f \in L_2(\Omega)$, $(\cdot,\cdot)$ is the $L_2(\Omega)$ inner product, and

$$a(w,v) = \sum_{i,j=1}^{2} \int_\Omega \frac{\partial^2 w}{\partial x_i \partial x_j} \frac{\partial^2 v}{\partial x_i \partial x_j}\, dx$$

is the inner product of the Hessian matrices of w and v.

Let p_* be a corner of Ω, and

$$\mathbb{V}^* = \{v \in \mathbb{V} : v(p_*) = 0\}.$$

Then by elliptic regularity [1], the unique solution $u \in V^*$ of our model problem belongs to $H^{2+\alpha}(\Omega)$, where $0 < \alpha \leq 2$ is the index of elliptic regularity.

R. Bank et al. (eds.), *Domain Decomposition Methods in Science and Engineering XX*, Lecture Notes in Computational Science and Engineering 91,
DOI 10.1007/978-3-642-35275-1_14,

C^0 interior penalty methods are discontinuous Galerkin methods for fourth order problems. These approaches for our model problem have recently been analyzed in [5]. Let $\mathscr{T}_h$ be a simplicial or convex quadrilateral triangulation of Ω, and V_h be a Lagrange (triangular or tensor product) finite element space associated with $\mathscr{T}_h$. Let

$$V_h^* = \{v \in V_h : v(p_*) = 0\}.$$

Then the C^0 interior penalty method for (1) and (2) is to find $u_h \in V_h^*$ such that

$$\mathscr{A}_h(u_h, v) = (f, v) \qquad \forall v \in V_h^*, \tag{3}$$

where for $w, v \in V_h^*$,

$$\begin{aligned}\mathscr{A}_h(w,v) = &\sum_{D\in\mathscr{T}_h}\sum_{i,j=1}^{2}\int_D \frac{\partial^2 w}{\partial x_i \partial x_j}\frac{\partial^2 v}{\partial x_i \partial x_j}\,dx + \sum_{e\in\mathscr{E}_h}\frac{\eta}{|e|}\int_e \left[\!\left[\frac{\partial w}{\partial n}\right]\!\right]\left[\!\left[\frac{\partial v}{\partial n}\right]\!\right] ds \\ &+ \sum_{e\in\mathscr{E}_h}\int_e \left(\left\{\!\!\left\{\frac{\partial^2 w}{\partial n^2}\right\}\!\!\right\}\left[\!\left[\frac{\partial v}{\partial n}\right]\!\right] + \left\{\!\!\left\{\frac{\partial^2 v}{\partial n^2}\right\}\!\!\right\}\left[\!\left[\frac{\partial w}{\partial n}\right]\!\right]\right) ds,\end{aligned} \tag{4}$$

$\mathscr{E}_h$ denotes the set of edges of the triangulation $\mathscr{T}_h$, and η is a penalty parameter. The jumps and averages are defined as follows.

For interior edges $e \in \mathscr{E}_h$ shared by two elements $D_\pm \in \mathscr{T}_h$, we take n_e to be the unit normal of e pointing from D_- into D_+, and define

$$\left[\!\left[\frac{\partial v}{\partial n}\right]\!\right] = \frac{\partial v_+}{\partial n_e} - \frac{\partial v_-}{\partial n_e} \quad \text{and} \quad \left\{\!\!\left\{\frac{\partial^2 v}{\partial n^2}\right\}\!\!\right\} = \frac{1}{2}\left(\frac{\partial^2 v_+}{\partial n_e^2} + \frac{\partial^2 v_-}{\partial n_e^2}\right),$$

where $v_\pm = v\big|_{D_\pm}$. Note that the definitions of $[\![\partial v/\partial n]\!]$ and $\{\!\{\partial^2 v/\partial n^2\}\!\}$ are independent of the choice of e.

For $e \in \mathscr{E}_h$ which is on the boundary of Ω, we take n_e to be the unit normal of e pointing outside Ω and define

$$\left[\!\left[\frac{\partial v}{\partial n}\right]\!\right] = -\frac{\partial v}{\partial n_e} \quad \text{and} \quad \left\{\!\!\left\{\frac{\partial^2 v}{\partial n^2}\right\}\!\!\right\} = \frac{\partial^2 v}{\partial n_e^2}.$$

Remark 1. The discrete problem (3) resulting from the C^0 interior penalty method is consistent, and for the penalty parameter η large enough, it is also stable [3].

For fourth order problems, C^0 interior penalty methods have certain advantages over classical finite element methods. However, due to the nature of fourth order problems, the discrete system resulting from the C^0 interior penalty method is very ill-conditioned. Therefore, it is necessary to develop modern fast solvers to overcome this drawback. In this paper, we construct a two-level additive Schwarz preconditioner and extend the results in [4] for biharmonic problems with essential Dirichlet boundary conditions to the ones with the essential and natural boundary conditions.

The rest of this paper is organized as follows. We first introduce the framework of a two-level additive Schwarz preconditioner in Sect. 2, followed by the condition number estimates of the preconditioned system in Sect. 3. Section 4 demonstrates some numerical results.

2 A Two-Level Additive Schwarz Preconditioner

For simplicity, we will focus on the case where $\mathscr{T}_h$ is a rectangular mesh. The results obtained in this paper are also true for triangular and general convex quadrilateral meshes.

Let $V_h^* = \{v : v \in C(\bar{\Omega}), v(p_*) = 0, v_D = v\big|_D = \mathbb{Q}_2(D)\ \forall D \in \mathscr{T}_h\}$ be the standard quadratic Lagrange finite element space associated with $\mathscr{T}_h$, and the operator $A_h : V_h^* \longrightarrow V_h^{*'}$ can then be defined by

$$\langle A_h v, w\rangle = \mathscr{A}_h(v, w) \qquad \forall v, w \in V_h^*,$$

where $\langle \cdot, \cdot \rangle$ is the canonical bilinear form between a vector space and its dual.

Note that for η sufficiently large, the following relation [3] is true.

$$C_1 |v|^2_{H^2(\Omega, \mathscr{T}_h)} \le \langle A_h v, v\rangle \le C_2 |v|^2_{H^2(\Omega, \mathscr{T}_h)} \qquad \forall v \in V_h^*,$$

where

$$|v|^2_{H^2(\Omega, \mathscr{T}_h)} = \sum_{D \in \mathscr{T}_h} |v|^2_{H^2(D)} + \sum_{e \in \mathscr{E}_h} \frac{1}{|e|} \|[\![\partial v/\partial n]\!]\|^2_{L_2(e)},$$

and the constants C_1 and C_2 depend only on the shape regularity of $\mathscr{T}_h$.

We now construct a two-level additive Schwarz preconditioner for the operator A_h which involves a coarse grid solve and subdomain solves.

First of all, let $\mathscr{T}_H$ be a coarse rectangular mesh for Ω, and $V_0 \subset H^1(\Omega)$ be the $\mathbb{Q}_1$ finite element space associated with $\mathscr{T}_H$. We define $A_0 : V_0^* \longrightarrow V_0^{*'}$ by

$$\langle A_0 v, w\rangle = \mathscr{A}_H(v, w) \qquad \forall v, w \in V_0^*,$$

where $\mathscr{A}_H$ is the analog of $\mathscr{A}_h$ for the coarse grid $\mathscr{T}_H$, and $V_0^* = \{v : v \in V_0, v(p_*) = 0\}$.

Let $\Omega_j, 1 \le j \le J$, be overlapping subdomains of Ω such that $\Omega = \cup_{j=1}^J \Omega_j$, and the boundaries of Ω_j are aligned with the edges of $\mathscr{T}_h$. We assume that there exist nonnegative $\theta_j \in C^\infty(\bar{\Omega})$ for $1 \le j \le J$ such that

$$\begin{aligned} \theta_j &= 0 \qquad \text{on} \quad \Omega \backslash \Omega_j, \\ \sum_{j=1}^J \theta_j &= 1 \qquad \text{on} \quad \bar{\Omega}, \\ \|\nabla \theta_j\|_{L_\infty(\Omega)} \le \frac{C}{\delta}, &\qquad \|\nabla^2 \theta_j\|_{L_\infty(\Omega)} \le \frac{C}{\delta^2}, \end{aligned}$$

where $\nabla^2 \theta_j$ is the Hessian of θ_j, $\delta > 0$ measures the overlap among the subdomains, and C is a positive constant independent of h, H and J.

Remark 2. Suppose $\mathscr{T}_h$ is a refinement of $\mathscr{T}_H$. We can construct Ω_j by enlarging the elements of $\mathscr{T}_H$ by the amount of δ so that the boundaries of $\Omega_j, 1 \le j \le J$, are aligned with the edges of $\mathscr{T}_h$ (cf. Fig. 1). The construction of $\theta_j, 1 \le j \le J$, is then standard.

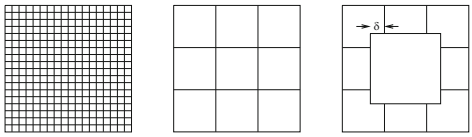


Fig. 1. $\mathscr{T}_h, \mathscr{T}_H$ and Ω_j

Moreover, we assume that the maximum number of subdomains Ω_j that share a common point is bounded by a constant N_c.

Let $V_j = \{v : v \in V_h^*, v = 0 \text{ on } \bar{\Omega}_\ell \text{ if } \ell \neq j\}$ be the $\mathbb{Q}_2$ finite element space associated with $\mathscr{T}_h$ on $\bar{\Omega}_j$. Then we define the operator $A_j : V_j \longrightarrow V_j'$ by

$$\langle A_j v, w\rangle = \mathscr{A}_j(v,w) \qquad \forall v, w \in V_j,$$

where $\mathscr{A}_j, 1 \leq j \leq J$, are the analogs of $\mathscr{A}_h$ restricted on $\bar{\Omega}_j$. Similarly, we obtain that

$$C_3|v|^2_{H^2(\Omega_j,\mathscr{T}_h)} \leq \langle A_j v, v\rangle \leq C_4|v|^2_{H^2(\Omega_j,\mathscr{T}_h)} \qquad \forall v \in V_j,$$

where

$$|v|^2_{H^2(\Omega_j,\mathscr{T}_h)} = \sum_{\substack{D\in\mathscr{T}_h \\ D\subset\bar{\Omega}_j}} |v|^2_{H^2(D)} + \sum_{\substack{e\in\mathscr{E}_h \\ e\subset\bar{\Omega}_j}} \|[\![\partial v/\partial n]\!]\|^2_{L_2(e)},$$

and C_3, C_4 are constants independent of h, H, J, N_c and δ.

For simplicity, from now on, we will use C to denote a generic positive constant independent of h, H, δ, and J that will take different values in different occurrences.

The subdomain finite element space $V_j, 1 \leq j \leq J$, is connected to V_h^* by the natural injection operator I_j which satisfies the following inequality.

$$|I_j v|_{H^2(\Omega,\mathscr{T}_h)} \leq C|v|_{H^2(\Omega_j,\mathscr{T}_h)} \qquad \forall v \in V_j.$$

Furthermore, the coarse space V_0^* and the fine space V_h^* are connected by the operator I_0 which is defined as follows.

Let $\tilde{V}_0 \subset H^2(\Omega)$ be the $\mathbb{Q}_3$ Bogner-Fox-Schmit finite element space associated with $\mathscr{T}_H$, and $\tilde{V}_0^* = \{v : v \in \tilde{V}_0, v(p_*) = 0\}$. The $\mathbb{Q}_1$ Lagrange element and the $\mathbb{Q}_3$ Bogner-Fox-Schmit element are depicted in Fig. 2, where we use the solid dot $\bullet$ to denote pointwise evaluation of the shape functions, the circle $\circ$ and the arrow $\nearrow$ to denote pointwise evaluation of all the first order derivatives and the mixed second order derivative of the shape functions, respectively.

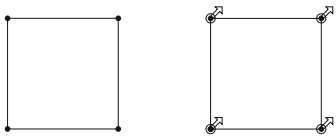


Fig. 2. $\mathbb{Q}_1$ element and $\mathbb{Q}_3$ Bogner-Fox-Schmit element

We define $E_H : V_0^* \longrightarrow \tilde{V}_0^*$ to be the operator that for all $p \in \mathscr{T}_H$,

$$
\begin{aligned}
(E_H v)(p) &= v(p), \\
\nabla(E_H v)(p) &= \begin{cases} \dfrac{1}{|\mathscr{T}_p|} \displaystyle\sum_{D \in \mathscr{T}_p} \nabla v_D(p), & \text{if } p \in \Omega, \\ 0, & \text{if } p \in \partial\Omega, \end{cases} \\
\frac{\partial^2 (E_H v)}{\partial x_1 \partial x_2}(p) &= \begin{cases} \dfrac{1}{|\mathscr{T}_p|} \displaystyle\sum_{D \in \mathscr{T}_p} \frac{\partial^2 v_D}{\partial x_1 \partial x_2}(p), & \text{if } p \in \Omega, \\ 0, & \text{if } p \in \partial\Omega, \end{cases}
\end{aligned}
$$

where $\mathscr{T}_p$ is the set of rectangles in $\mathscr{T}_H$ sharing p as a vertex, $|\mathscr{T}_p|$ is the number of elements in $\mathscr{T}_p$, and $v_D = v\big|_D$.

Then for all $v \in V_0^*$, we take $I_0 v \in V_h^*$ to be the one whose nodal values are identical with the corresponding nodal values of $E_H v$.

Remark 3. Instead of using the operator E_H, if we define the operator I_0 as the natural injection operator from V_0^* to V_h^*, then the performance of the preconditioner will be affected by the different scalings that appear in the penalty terms for $\mathscr{A}_h$ and $\mathscr{A}_H$. However, this problems can be avoided by defining I_0 as above since $E_H v \in H^2(\Omega)$.

We can now define the two-level additive Schwarz preconditioner $B : V_h^{*'} \longrightarrow V_h^*$ by

$$B = \sum_{j=0}^{J} I_j A_j^{-1} I_j^t,$$

where $I_j^t : V_h^{*'} \longrightarrow V_j'$ is the transpose of I_j, i.e.,

$$\langle I_j^t \Psi, v \rangle = \langle \Psi, I_j v \rangle \qquad \forall \Psi \in V_h^{*'}, v \in V_j.$$

From the additive Schwarz theory [2, 6], the preconditioner B is symmetric positive definite and therefore the eigenvalues of BA_h are positive. Moreover, the maximum and minimum eigenvalues of BA_h are given by the following formulas, which will be used to estimate the condition number of the preconditioned system.

$$
\begin{aligned}
\lambda_{\max}(BA_h) &= \max_{\substack{v \in V_h \\ v \neq 0}} \frac{\langle A_h v, v \rangle}{\min\limits_{\substack{v = \sum_{j=0}^{J} I_j v_j \\ v_j \in V_j}} \sum\limits_{j=0}^{J} \langle A_j v_j, v_j \rangle}, \\
\lambda_{\min}(BA_h) &= \min_{\substack{v \in V_h \\ v \neq 0}} \frac{\langle A_h v, v \rangle}{\min\limits_{\substack{v = \sum_{j=0}^{J} I_j v_j \\ v_j \in V_j}} \sum\limits_{j=0}^{J} \langle A_j v_j, v_j \rangle}.
\end{aligned}
$$

3 Condition Number Estimates

From the construction of our two-level additive Schwarz preconditioner, by the similar arguments as we did in [4], it is not difficult to derive the following results on the estimates of the eigenvalues of the preconditioned system.

Theorem 1. *The following upper bound for the eigenvalues of BA_h holds:*

$$\lambda_{\max}(BA_h) \leq C,$$

where the positive constant C depends on the shape regularity of $\mathscr{T}_h$ and $\mathscr{T}_H$ but not h, H, δ nor J.

Theorem 2. *The following lower bound for the eigenvalues of BA_h holds:*

$$\lambda_{\min}(BA_h) \geq C\left(1+\frac{H^4}{\delta^4}\right),$$

where the positive constant C depends on the shape regularity of $\mathscr{T}_h$ and $\mathscr{T}_H$ but not h, H, δ nor J.

Finally, from Theorems 1 and 2, the following estimate on the condition number of the preconditioned system can be obtained immediately.

Theorem 3. *It holds that*

$$\kappa(BA_h) = \frac{\lambda_{\max}(BA_h)}{\lambda_{\min}(BA_h)} \leq C\left(1+\frac{H^4}{\delta^4}\right),$$

where the positive constant C depends on the shape regularity of $\mathscr{T}_h$ and $\mathscr{T}_H$ but not h, H, δ nor J.

Remark 4. In the case of a small overlap, i.e. $\delta \ll H$, the estimate on the condition number of the preconditioned system can be improved to $(1+(H/\delta)^3)$, provided with more assumptions on the subdomains Ω_j [4].

4 Numerical Results

In this section, we present some numerical results for the biharmonic problem with Cahn-Hilliard type of boundary conditions on the unit square. We choose the penalty parameter in $\mathscr{A}_h, \mathscr{A}_H$ and $\mathscr{A}_j$ to be 5, which guarantees the coerciveness of the variational form (4) on V_h^*.

First of all, for different choices of H and h, we generate a vector $v_h \in V_h^*$, compute the right-hand side vector $g = A_h v_h$, and apply the preconditioned conjugate gradient algorithm to the system $A_h z = g$ using our two-level additive Schwarz preconditioner. We compute the iteration numbers needed for reducing the energy norm error by a factor of 10^{-6} for five random choices of v_h and then average them. The

numbers are collected in Tables 1 and 2. Also, to illustrate the practical performance of our preconditioner, such iteration numbers needed for reducing the energy norm error by a factor of 10^{-2} with 16 subdomains are reported in Table 3. They show that the bound for the condition number of BA_h is independent of h.

We also compute, in the case of 4 and 16 subdomains, the maximum eigenvalue, the minimum eigenvalue, and the condition number of the preconditioned system for the fine mesh $h = 2^{-6}$ and various overlaps among subdomains by using Lanczos methods. The results are tabulated in Tables 4 and 5. They show that the maximum eigenvalue is bounded and the minimum eigenvalue increases as the overlap among subdomains decreases.

Table 1. Average number of iterations for reducing the energy norm error by a factor of 10^{-6} with $H = 1/2$ and $J = 4$

	$h = 2^{-2}$	$h = 2^{-3}$	$h = 2^{-4}$	$h = 2^{-5}$	$h = 2^{-6}$
$\delta = 2^{-2}$	17	17	17	15	15
$\delta = 2^{-3}$	-	20	20	19	17
$\delta = 2^{-4}$	-	-	26	25	24
$\delta = 2^{-5}$	-	-	-	47	45
$\delta = 2^{-6}$	-	-	-	-	93

Table 2. Average number of iterations for reducing the energy norm error by a factor of 10^{-6} with $H = 1/4$ and $J = 16$

	$h = 2^{-3}$	$h = 2^{-4}$	$h = 2^{-5}$	$h = 2^{-6}$
$\delta = 2^{-3}$	27	29	27	24
$\delta = 2^{-4}$	-	28	26	24
$\delta = 2^{-5}$	-	-	42	39
$\delta = 2^{-6}$	-	-	-	83

Bibliography

[1] H. Blum and R. Rannacher. On the boundary value problem of the biharmonic operator on domains with angular corners. *Math. Methods Appl. Sci.*, 2:556–581, 1980.

[2] S.C. Brenner and L.R. Scott. *The Mathematical Theory of Finite Element Methods*. Springer-Verlag, New York, third edition, 2008.

Table 3. Average number of iterations for reducing the energy norm error by a factor of 10^{-2} with $H = 1/4$ and $J = 16$

	$h = 2^{-3}$	$h = 2^{-4}$	$h = 2^{-5}$	$h = 2^{-6}$
$\delta = 2^{-3}$	6	6	5	5
$\delta = 2^{-4}$	-	5	5	4
$\delta = 2^{-5}$	-	-	5	4
$\delta = 2^{-6}$	-	-	-	5

Table 4. $\lambda_{\max}(BA_h), \lambda_{\min}(BA_h)$ and $\kappa(BA_h)$ with $H = 1/2, h = 2^{-6}$ and $J = 4$

H/δ	$\lambda_{\max}(BA_h)$	$\lambda_{\min}(BA_h)$	$\kappa(BA_h)$
2	4.8394	0.4259	1.1363×10^1
4	4.8029	0.3045	1.5775×10^1
8	4.7526	0.1279	3.7149×10^1
16	4.6600	0.0247	1.8850×10^2
32	4.5849	0.0036	1.2895×10^3

Table 5. $\lambda_{\max}(BA_h), \lambda_{\min}(BA_h)$ and $\kappa(BA_h)$ with $H = 1/4, h = 2^{-6}$ and $J = 16$

H/δ	$\lambda_{\max}(BA_h)$	$\lambda_{\min}(BA_h)$	$\kappa(BA_h)$
2	6.5195	0.1811	3.5992×10^1
4	4.8740	0.1633	2.9852×10^1
8	4.6968	0.0631	7.4402×10^1
16	4.5865	0.0103	4.4698×10^2

[3] S.C. Brenner and L.-Y. Sung. C^0 interior penalty methods for fourth order elliptic boundary value problems on polygonal domains. *J. Sci. Comput.*, 22/23:83–118, 2005.

[4] S.C. Brenner and K. Wang. Two-level additive Schwarz preconditioners for C^0 interior penalty methods. *Numer. Math.*, 102:231–255, 2005.

[5] S.C. Brenner, S. Gu, T. Gudi, and L.-Y. Sung. A C^0 interior penalty method for a biharmonic problem with essential and natural boundary conditions of Cahn-Hilliard type. 2010.

[6] M. Dryja and O.B. Widlund. An additive variant of the Schwarz alternating method in the case of many subregions. Technical Report 339, Department of Computer Science, Courant Institute, 1987.

An Algebraic Multigrid Method Based on Matching in Graphs

James Brannick[1], Yao Chen[1], Johannes Kraus[2], and Ludmil Zikatanov[1]

[1] Department of Mathematics, The Pennsylvania State University, PA 16802, USA;
Email: {brannick|chen_y|ltz}@math.psu.edu.

[2] Radon Institute for Computational and Applied Mathematics, Austrian Academy of Sciences, Altenbergerstraße 69, A-4040 Linz, Austria;
Email: johannes.kraus@oeaw.ac.at.

1 Introduction

We present an Algebraic Multigrid (AMG) method for graph Laplacian problems. The coarse graphs are constructed recursively by pair-wise aggregation, or matching as in [3] and we use an Algebraic Multilevel Iterations (AMLI) [1, 6] for the solution phase.

The two-level method constructs a splitting of the underlying vector space into two subspaces V_S and V_P and then corrects the error successively on V_S and V_P. The coarse space V_P is obtained using matching on the underlying graph. Such a two-level method is shown to be uniformly convergent. In the AMLI method (multilevel), m coarse level corrections are applied on each level. For large m, while the convergence rate of the method is comparable to that of the two-level method and, hence, uniformly convergent, it is clear that the overall complexity of such method could be too high for large values of m. In our approach, the AMLI convergence rate is estimated solely based on the underlying two-level method, which allows us to show that $m = 2$ gives a balance between the complexity and the desired convergence rate, thus, resulting in an efficient algorithm.

The paper is organized as follows. In Sect. 2 the graph Laplacian problem is described. In Sect. 3, the graph matching algorithm is introduced and it is indicated that the ℓ_2 projection on the coarse space is the key quantity for obtaining the multilevel estimates of the AMLI method. In Sect. 4, an analysis of a specific two-level method is presented and in Sect. 5 its convergence and complexity are estimated. In the following section, numerical results are reported.

2 Graph Laplacian Problems

Graph Laplacian solvers can be used as preconditioners for various discrete numerical models, e.g., ones arising from discretizations of partial differential equations,

R. Bank et al. (eds.), *Domain Decomposition Methods in Science and Engineering XX*, Lecture Notes in Computational Science and Engineering 91,
DOI 10.1007/978-3-642-35275-1_15,

machine learning algorithms, and spectral clustering of images. Consider a connected unweighted graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ where $\mathcal{V}$ and $\mathcal{E}$ are the sets of vertices and edges. The graph Laplacian $A \in \mathbb{R}^{n\times n}$, where $n = |\mathcal{V}|$ (cardinality of $\mathcal{V}$), corresponding to the graph $\mathcal{G}$, can be defined as follows:

$$(Au, v) = \sum_{k=(i,j)\in\mathcal{E}} (u_i - u_j)(v_i - v_j).$$

The matrix A is symmetric and positive semi-definite. The null space of A is one dimensional, and its basis is given by $\{\mathbf{1}\}$, where $\mathbf{1}$ is a vector whose components are all equal to 1. Our aim here is to solve graph Laplacian problems, or to find u, such that $(u, \mathbf{1}) = 0$ and

$$Au = f,$$

for a given f satisfying $(f, \mathbf{1}) = 0$.

We want to find an AMG method to solve graph Laplacians with simple settings, so that we can estimate the performance of the AMG method, with as few assumptions introduced as possible. The construction of this AMG method can also help us to derive similar methods for weighted graph Laplacian problems, which come from finite element or finite difference discretizations of elliptic partial differential equations, circuit simulations, and in general, network flow simulations.

3 Graph Matching

Given a graph $\mathcal{G}$, assume that we can find a set of aggregates $\mathcal{M}$ called a *matching*, where each aggregate contains exactly two vertices, and every vertex of $\mathcal{G}$ is contained in exactly one aggregate. For a certain aggregate that contains vertices i and j, we merge the two vertices, and the newly formed vertex, named k, is considered connected to the vertex l if and only if l is connected to i or j on graph $\mathcal{G}$. By merging vertices in each aggregate, a reduced graph of the graph $\mathcal{G}$ is formed. Applying such a matching algorithm recursively will result in a sequence of graphs. We then construct a solver for the graph Laplacian of $\mathcal{G}$ based on the sequence of reduced graphs.

In the matching $\mathcal{M}$, we consider the k-th aggregate as a graph $\mathcal{G}_k = (\mathcal{V}_k, \mathcal{E}_k)$. Let Q be the ℓ_2-orthogonal projection on the coarse space, which consists of vectors that are piecewise constant on each set $\mathcal{V}_k$. An alternative definition of Q is as follows.

$$(Qu)_i = \frac{1}{|\mathcal{V}_k|} \sum_{j\in\mathcal{V}_k} u_j, \quad i \in \mathcal{V}_k.$$

Classical AMG theory suggests that the coarse space should cover, or approximate algebraically smooth error components. Detailed explanations can be found, e.g., in the appendix of [5]. In the following section, we will compute how well piecewise constant vectors can approximate smooth vectors and will discuss the properties of two-level and multilevel methods using the subspace(s) associated with the projection Q.

4 A Two-Level Method

Define matrices P and S for a given matching $\mathscr{M}$, such that

$$P \cdot e_k = e_i + e_j, \quad S \cdot e_k = e_i - e_j, \quad (i,j) \in \mathscr{V}_k,$$

where e_i and e_j are Euclidean basis vectors. Since a prerequisite for designing an efficient AMLI method is an efficient two-level method, in this section we focus on two-level methods and their convergence rates. Given an initial guess u_0, a typical two-level algorithm which takes as input u_k and returns the next iterate u_{k+1} is as follows:

1. $v = u_k + SR^{-1}S^T(f - Au_k)$,
2. $w = v + PA_c^{-1}P^T(f - Av)$,
3. $u_{k+1} = w + SR^{-T}S^T(f - Aw)$.

Here the matrix R is a preconditioner of S^TAS, which is the restriction of A on the space $\text{range}(S) = [\text{range}(P)]^{\perp}$. The matrix A_c is an approximation of the restriction of A on the coarse space $V_c = \text{range}(P)$. In our algorithm, A_c is first defined as the graph Laplacian of the unweighted coarse graph and thus $A_c \neq P^TAP$. We then scale A_c such that $(v^TA_cv)/(v^TP^TAPv) \in [1, c_c]$. A proper scaling results in $c_c = 2$ for P that corresponds to an aligned matching and A that is a structured grid of any dimension. The matrix representation of this two-level method, denoted by G, can be deduced via the error propagation matrix given as follows.

$$E = (I - SR^{-T}S^TA)(I - PA_c^{-1}P^TA)(I - SR^{-1}S^TA) = I - G^{-1}A. \tag{1}$$

We now derive an estimate on the angle between the spaces $\text{range}(S)$ and $\text{range}(P)$, which in our setting amounts to obtaining a bound on the energy norm of Q, the ℓ_2-orthogonal projection onto $\text{range}(P)$. Let γ be the C.B.S. constant such that it is the smallest number satisfying $(Sw, Pv)_A \leq \gamma |Sw|_A |Pv|_A$, then (cf. [6, Corollary 3.7]):

$$|Q|_A^2 = 1/(1 - \gamma^2).$$

Using [2, Theorem 4.2] we can show that, if the symmetrized smoother $\widetilde{R} = R + R^T - S^TAS$ is positive definite, and $(w^T\widetilde{R}w)/(w^TS^TASw) \in [1, \kappa_s]$, then

$$\frac{v^TGv}{v^TAv} \in [1, |Q|_A^2(\kappa_s + c_c - 1)].$$

If a two-level method using a certain matching is already given, then both $|Q|_A$ and κ_s can be estimated using the properties of the underlying graph. The norm $|Q|_A$ is estimated as follows:

$$u^TQAQu = \sum_{(i,j)\in\mathscr{E}} ((Qu)_i - (Qu)_j)^2 \leq 2d \sum_{(i,j)\in E} (u_i - u_j)^2 \leq (2d)u^TAu$$

where d is the maximum degree of the graph. This implies that $|Q|_A^2 \leq 2d$. Assuming that the matching $\mathscr{M}$ is perfect, we show that the smallest eigenvalue of S^TAS is larger or equal to 4, by computing

$$w^T S^T ASw \geq \sum_{(i,j)\in\mathscr{M}} ((Sw)_i - (Sw)_j)^2 = \sum_{(i,j)\in\mathscr{M}} 4(Sw)_i^2 = 4\|w\|_{\ell_2}^2.$$

According to the Gershgorin theorem, the largest eigenvalue of $S^T AS$ is bounded by a function of d and for a simple smoother R, such as Richardson iteration, κ_s is also bounded by a function of d. From the above results (i.e, the stability estimate of Q in the A-seminorm and the lower bound on the smallest eigenvalue of $S^T AS$) it follows that the two-level method is uniformly convergent with respect to the size of the matrix A. Based on the two-level convergence estimate, AMLI cycles with low complexity and predictable convergence is then constructed.

5 Algebraic Multilevel Iterations

An estimate of the two-level convergence rate does not automatically carry over to an estimate of the convergence of a multilevel V-cycle, and in general, for piece-wise constant coarse spaces, it can be shown that the convergence rate degrades exponentially with respect to the number of levels. A remedy for this issue is to use more complicated cycles such as AMLI, and keep a balance between complexity of a cycle and its convergence rate so that the resulting algorithm is optimal or nearly optimal.

We describe an AMLI method by first rewriting the two-level preconditioner G, as well as $\widehat{G}$ which is G under the hierarchical basis (S,P), in block form:

$$\begin{aligned}\widehat{G}^{-1} &= \widehat{L}^{-T}\begin{pmatrix}(R+R^T-S^TAS)^{-1} & 0\\ 0 & A_c^{-1}\end{pmatrix}\widehat{L}^{-1},\\ G &= (S,P)^{-1}\widehat{G}(S,P)^{-T},\end{aligned}$$

where

$$\widehat{L} = \begin{pmatrix} I & 0\\ P^T ASR^{-1} & I\end{pmatrix}.$$

Then define an AMLI preconditioner B as follows.

$$\begin{aligned}\widehat{B}^{-1} &= \widehat{L}^{-T}\begin{pmatrix}(R+R^T-S^TAS)^{-1} & 0\\ 0 & B_c^{-1}q(A_cB_c^{-1})\end{pmatrix}\widehat{L}^{-1},\\ B^{-1} &= (S,P)^T\widehat{B}^{-1}(S,P).\end{aligned}$$

Here A_c is the scaled unweighted graph Laplacian of the coarse graph and B_c is a preconditioner of A_c, and $q(t)$ is a polynomial. When $q(t)=1$, the action $\widehat{B}^{-1}$ stands for a V-cycle with an inexact solver B_c^{-1} on the coarse level. In the case of a W-cycle, we have $q(t)=2-t$.

The following lemma shows how well the AMLI preconditioner B approximates the two-level preconditioner G.

Lemma 1. *If $\lambda_1 \leq \lambda(B_c^{-1}A_c) \leq \lambda_2$ and $tq(t) > 0$ for $t \in [\lambda_1,\lambda_2]$, then*

$$\min(1, \min_{\lambda_1\leq t\leq\lambda_2}\frac{1}{tq(t)}) \leq \frac{v^T G^{-1}v}{v^T B^{-1}v} \leq \max(1, \max_{\lambda_1\leq t\leq\lambda_2}\frac{1}{tq(t)}).$$

This lemma suggests that, the AMLI method is spectrally equivalent to a two-level method, given that the coarse-level preconditioner is spectrally equivalent to the coarser-level matrix. The upper and lower bounds in the lemma above are related to estimates on $|tq(t)|$ for t in a given interval. As shown in [1, 6], using higher order polynomials $q(t)$, the matrix B^{-1} can approximate G^{-1} arbitrarily well and thus we will have a method with excellent convergence rate. However, a higher order polynomial $q(t)$ leads to a much more expensive computation of the coarser level correction, and the resulting multilevel methods can have a very high complexity and one should be careful in the choice of the polynomial degree.

Assume that a multilevel hierarchy is formed by a recursive application of the matching algorithm. Denote the graph Laplacians on each level, and the corresponding two-level preconditioners by A_k and G_k. Following the ordering of levels in [1, 6] we set $A = A_0$ and denote by A_J the coarsest matrix. Define a sequence of solvers as

$$\begin{aligned}
\widehat{B}_J^{-1} &= \widehat{A}_J^{\dagger} = (S_J, P_J)^{-T} A_J^{\dagger} (S_J, P_J)^{-1}, \\
B_k^{-1} &= (S_k, P_k)^T \widehat{B}_k^{-1} (S_k, P_k), \quad k = 0, \ldots, J, \\
\widehat{B}_k^{-1} &= \widehat{L}_k^{-T} \begin{pmatrix} (R_k + R_k^T - S_k^T A_k S_k)^{-1} & 0 \\ 0 & B_{k+1}^{-1} q(A_{k+1} B_{k+1}^{-1}) \end{pmatrix} L_k^{-1}, \quad k = 0 \ldots J-1.
\end{aligned}$$

Then, a multilevel proof of convergence follows.

Lemma 2. *Assume that there is a constant c_g, $1 \le c_g < 4$, such that the following relation holds.*

$$v^T \widehat{A}_k v \le v^T \widehat{G}_k v \le c_g v^T \widehat{A}_k v, \quad \forall v \text{ and } k = 0, \ldots, J.$$

Then there exists a linear function $q(t)$, such that

$$\frac{2}{\sqrt{c_g}} - 1 \le \frac{v^T B_k^{-1} v}{v^T A_k^{-1} v} \le 1, \quad \forall v \text{ and } k = 0, \ldots, J.$$

Here $q(t)$ is a scaled and shifted Chebyshev type polynomial (see [1]).

This lemma shows that, if c_g is strictly less than 4, then the action B_0^{-1} is an uniformly convergent AMLI cycle with $O(n \log n)$ complexity. Even if $c_g = 4$ on all levels, one may prove that the condition number of $B_J^{-1} A_J$ for the case of second order $q(t)$ (similar to a W-cycle) grows linearly with respect to the number of levels $J = \log n$. This results in a convergence factor $1 - 1/\log n$ at a complexity of $O(n \log n)$ for each cycle.

The two-level method we suggest is based on graph matching, thus $c_g \le |Q|_A^2 (\kappa_s + c_c - 1)$. In a simple case where the graph $\mathcal{G}$ is a two-dimensional uniform grid, an aligned regular matching yields $|Q|_A^2 \le 2$, $\kappa_s = 1 + \varepsilon$ for arbitrary small ε, and $c_c \le 2$.

This yields $c_g \le 4$ and thus the W-cycle AMLI preconditioner will result in a nearly optimal order method (cf. Lemma 2 and the discussion below). For unstructured or higher dimensional grids, numerical experiments indicate that random matching may still result in two-level methods for which $c_g \le 4$.

6 Numerical Results

We use the matching based AMLI method to solve a family of unweighted graph Laplacians, corresponding to graphs that represent structured grids or unstructured triangulations.

Structured grids. In the structured grid case on a rectangular domain, we match in a fixed direction. After several levels of matching the graph corresponding to the coarsest grid is a line. For the test on L-shaped domain, we still use matching in a fixed direction until a part of the coarsest graph becomes a tree. In such case, the unknowns can be ordered so that the fill-in during LU factorization on the coarsest grid is small.

A similar strategy can be used for graph Laplacians corresponding to three-dimensional structured grids. The matching procedure is applied only in two fixed directions.

Convergence analysis indicates that, choosing as a smoother $R^{-1} = (S^T AS)^{-1}$ guarantees the bound $c_g \leq 4$, for a matching based two-level method on structured grids. In the numerical experiments, we instead use a Gauss-Seidel smoother for all structured grid problems. Using such a smoother retains a convergence rate $\sim(1 - 1/\log n)$ and $O(n \log n)$ computational complexity.

Unstructured grids. Each of the unstructured grids in our tests are constructed by first perturbing the coordinates of vertices of a structured grid, followed by Delaunay triangulation of the resulting set of vertices. For unstructured grids, we use a random matching algorithm. Numerical results show that the maximum degree of the coarser graphs grow only during the first few coarsening steps. Hence, smoothers such as Gauss-Seidel can approximate well $(S_k^T A_k S_k)^{-1}$ on all levels and the application of such a smoother has a complexity proportional to the number of degrees of freedom (DOF) on level k. We use the CG method to perform the action of $(S_k^T A_k S_k)^{-1}$ on a vector. Such approach is practical since $S^T AS$ is equally well conditioned on all levels.

Instead of using the same AMLI polynomial $q(t)$ on all levels, we determine the polynomials $q_k(t)$ on each level recursively, starting from the second coarsest level. After constructing a multilevel hierarchy, we use 6 AMLI two level cycles (level $(J-1)$ and level J) and a Lanczos algorithm to estimate the condition number of $B_{J-1}^{-1} A_{J-1}$. We apply this procedure recursively (and with 6 AMLI *multilevel* cycles from level $(k+1)$ to J) to estimate the condition number of $B_k^{-1} A_k$ on level k, for $k = 1, \ldots, J-2$. When all polynomials are determined, they are used in the AMLI cycle during the solving phase.

Numerical tests. We use the AMLI cycle as a preconditioner of Conjugate Gradient (CG) method. We stop the iterations when the relative residual becomes smaller than 10^{-10}. The results are summarized in Table 1. The number of CG iterations is denoted by M, and the average convergence rate of the last five iterations is denoted by r_a. The CG coefficients are also used to estimate the condition number $\kappa(B_0^{-1} A_0)$, as suggested in [4]. The operator and grid complexities are less than 2 in all the examples presented below.

(a) 2D unit square

DOF	κ	r_a	M
256^2	18.4	0.55	32
512^2	24.8	0.61	36
1024^2	32.9	0.69	40

(b) 3D unit cube

DOF	κ	r_a	M
32^3	7.8	0.36	21
64^3	11.4	0.45	25
128^3	19.2	0.51	29

(c) 2D L-Shaped

DOF	κ	r_a	M
$(3/4)\cdot 256^2$	17.8	0.56	33
$(3/4)\cdot 512^2$	23.9	0.64	36
$(3/4)\cdot 1024^2$	31.7	0.69	38

(d) 3D Fichera

DOF	κ	r_a	M
$(7/8)\cdot 32^3$	7.5	0.40	22
$(7/8)\cdot 64^3$	11.1	0.48	25
$(7/8)\cdot 128^3$	15.8	0.55	29

(e) 2D unit square (ug)

DOF	κ	r_a	M
256^2	31.4	0.58	35
512^2	36.7	0.63	39
1024^2	42.0	0.58	41

(f) 3D unit cube (ug)

DOF	κ	r_a	M
32^3	29.5	0.51	35
64^3	37.6	0.68	46
128^3	48.3	0.72	52

Table 1. Results for structured grids on square, cubic, L-shaped and Fichera domain, and for unstructured grids (ug) on square and cubic domain. Here, κ is an estimate (from CG) of $\kappa(B_0^{-1}A_0)$.

Note that for the 2D and 3D unstructured grid problems, the number of levels for a given unstructured grid is the same as that of a structured grid with the same degrees of freedom. We observe a logarithmic growth of the condition numbers with respect to the size of the grids, and fast convergence rates of the preconditioned CG method in all cases.

7 Conclusions

We present an AMLI (AMG) method based on graph matching with a nearly optimal convergence rate and computational complexity. We have also presented numerical tests which confirming our estimates. Our ongoing research is on extending the estimates to general aggregation algorithms and aggregates configurations and we are also investigating improvements of the AMLI method components.

Acknowledgments The authors gratefully acknowledge the support by the Austrian Academy of Sciences and by the Austrian Science Fund (FWF), Project No. P19170-N18 and the support from the National Science Foundation under grants NSF-DMS 0810982 and NSF-OCI 0749202.

Bibliography

[1] O. Axelsson and P. S. Vassilevski. Algebraic multilevel preconditioning methods. II. *SIAM J. Numer. Anal.*, 27(6):1569–1590, 1990.

[2] R. D. Falgout, P.S. Vassilevski, and L.T. Zikatanov. On two-grid convergence estimates. *Numer. Linear Algebra Appl.*, 12(5–6):471–494, 2005.

[3] H. Kim, J. Xu, and L. Zikatanov. A multigrid method based on graph matching for convection-diffusion equations. *Numer. Linear Algebra Appl.*, 10(1–2):181–195, 2003.

[4] Yousef Saad. *Iterative methods for sparse linear systems*. SIAM, Philadelphia, PA, 2nd edition, 2003.

[5] U. Trottenberg, C. W. Oosterlee, and A. Schüller. *Multigrid*. Academic Press Inc., San Diego, CA, 2001.

[6] P. S. Vassilevski. *Multilevel block factorization preconditioners*. Springer, New York, 2008.

Shifted Laplacian RAS Solvers for the Helmholtz Equation

Jung-Han Kimn[1] and Marcus Sarkis[2]

[1] Department of Mathematics and Statistics, South Dakota State University, Brookings, SD 57007, USA jung-han.kimn@sdstate.edu

[2] Mathematical Sciences Department, Worcester Polytechnic Institute, 100 Institute Road, Worcester, MA 01609, USA msarkis@wpi.edu and Instituto de Matemática Pura e Aplicada (IMPA), Brazil.

1 Introduction

We consider the Helmholtz equation:

$$-\Delta u^* - k^2 u^* = f \quad \text{in} \quad \Omega \tag{1}$$
$$u^* = g_D \text{ on } \partial\Omega_D, \ \frac{\partial u^*}{\partial n} = g_N \text{ on } \partial\Omega_N, \ \frac{\partial u^*}{\partial n} + iku^* = g_S \text{ on } \partial\Omega_S$$

where Ω is a bounded polygonal region in $\Re^2$, and the $\partial\Omega_D$, $\partial\Omega_N$ and $\partial\Omega_S$ correspond to subsets of $\partial\Omega$ where the Dirichlet, Neumann and Sommerfeld boundary conditions are imposed.

The main purpose of this paper is to introduce novel two-level overlapping Schwarz methods for solving the Helmholtz equation. Among the most effective parallel two-level domain decomposition solvers for the Helmholtz equation on general unstructured meshes, we mention the FETI-H method introduced by Farhat et al. [5], and the WRAS-H-RC method introduced by Kimn and Sarkis [10]. FETI-H type preconditioners belong to the class of nonoverlapping domain decomposition methods. FETI-H methods can be viewed as a modification of the original FETI method introduced by Farhat et al. [6]. The local solvers in FETI-H are based on Sommerfeld boundary conditions, see [3], while the coarse problem is based on plane waves. WRAS-H-RC type preconditioners belong to the class of overlapping Schwarz methods. They can be viewed as a miscellaneous of several methods to enhance the effectiveness of the solver for Helmholtz problems. The first ingredient of WRAS-H-RC preconditioners is the use of Sommerfeld boundary conditions for the local solvers on overlapping subdomains. This idea is similar to what was done in FETI-H, however, now for the overlapping case. This idea can be found for instance in the work of Cai et al. [2] and Kimn [8]. The second ingredient is the use of the Weighted Restricted Additive Schwarz (WRAS) method introduced by Cai and Sarkis [1] in order to average the local overlapping solutions. The third ingredient is the use of

R. Bank et al. (eds.), *Domain Decomposition Methods in Science and Engineering XX*, Lecture Notes in Computational Science and Engineering 91,
DOI 10.1007/978-3-642-35275-1_16,

partition of unity coarse spaces, see [13]. Here we consider the multiplication of a partition of unity times plane waves; see [12]. The fourth ingredient is how to define the coarse problem. It was discovered in [10] that a dramatic gain in performance can be obtained if WRAS techniques are applied to the fine-to-coarse restriction operator and the coarse-to-fine prolongation operator. The idea is to force the to act more locally on the fine-to-coarse transference of information and globally on the coarse-to-fine phase. The last ingredient is to put all these pieces together. The idea is to extend the Balancing Domain Decomposition (BDD) methods of Mandel [11], which were originally developed for the nonoverlapping case, to the overlapping case. This extension was introduced in [9] and the methods there were denoted by Overlapping Balancing Domain Decomposition (OBDD) methods. The WRAS-H-RC methods in [10] stand for "WRAS" for the local solvers, "H" for the FETI-H ingredients included in the methods, and "RC" for the restricted flavor of coarse problem.

Here in this paper we investigate numerically new techniques to improve further the performance of the WRAS-H-RC. More precisely, the shifted Laplacian techniques introduced in [7] and [4], are used to construct novel local solvers. We investigate how the various kinds of shifts affect the performance of the algorithms. As a result, we discover novel preconditioners that are more effective than the existing ones.

2 Discrete Formulation of the Problem

From a Green's formula, (1) can be reduced to: Find $u^* - u_D^* \in H_D^1(\Omega)$ such that,

$$\begin{aligned} a(u^*, v) &= \int_\Omega (\nabla u^* \cdot \nabla \bar{v} - k^2 u^* \bar{v})\, dx + ik \int_{\partial\Omega_S} u^* \bar{v}\, ds \\ &= \int_\Omega f \bar{v}\, dx + \int_{\partial\Omega_N} g_N \bar{v}\, ds + \int_{\partial\Omega_S} g_S \bar{v} = F(v),\ \forall v \in H_D^1(\Omega), \end{aligned} \tag{2}$$

where u_D^* is an extension of g_D to $H^1(\Omega)$, and $H_D^1(\Omega)$ is the space of $H^1(\Omega)$ functions vanishing on $\partial\Omega_D$.

Let $\mathscr{T}_h(\Omega)$ be a quasi-uniform triangulation of Ω and let $V \subset H_D^1(\Omega)$ be the finite element space of continuous piecewise linear functions vanishing on $\partial\Omega_D$. We assume that g_D on $\partial\Omega_D$ is a piecewise linear continuous function on $\mathscr{T}^h(\partial\Omega_D)$ and we have eliminated g_D by a discrete trivial zero extension inside Ω. We then obtain a discrete problem of the following form: Find $u \in V$ such that

$$a(u, v) = f(v), \quad \forall\, v \in V. \tag{3}$$

Using the standard hat basis functions, (3) can be rewritten as a linear system of equations of the form

$$Au = f. \tag{4}$$

3 Description of the WRAS-H-RC Methods

3.1 Partitioning and Subdomains

Given the triangulation $\mathscr{T}^h(\Omega)$, we assume that a domain partition by elements has been applied and resulted in N nonoverlapping subdomains $\Omega_i, i = 1, \ldots N$, such that

$$\overline{\Omega} = \cup_{i=1}^{N} \overline{\Omega}_i \text{ and } \Omega_i \cap \Omega_j = \emptyset, \text{ for } j \neq i.$$

Let δ be a nonnegative integer. Define $\Omega_i^0 = \Omega_i$. For $\delta \geq 1$, define the overlapping subdomains Ω_i^δ as follows: let Ω_i^1 be the one-overlap element extension of Ω_i^0 by including all the immediate neighboring elements $\tau_h \in \mathscr{T}^h(\Omega)$ such that $\overline{\tau}_h \cap \overline{\Omega}_i^0 \neq \emptyset$. Using this idea recursively, we can define a δ-extension overlapping subdomains Ω_i^δ

$$\Omega_i = \Omega_i^0 \subset \Omega_i^1 \subset \cdots \subset \Omega_i^\delta \cdots$$

3.2 Partition of the Unity

Let w be a nonnegative integer. For nodes x on $\partial\Omega_i^0$ define $\hat{\vartheta}_i^w(x) = 1$, for nodes x on $\partial\Omega_i^1 \backslash \overline{\Omega}_i^0$ define $\hat{\vartheta}_i^w(x) = 1 - 1/(w+1)$, for nodes x on $\partial\Omega_i^2 \backslash \overline{\Omega}_i^1$ define $\hat{\vartheta}_i^w(x) = 1 - 2/(w+1)$, and recursively until $\hat{\vartheta}_i^w(x) = 0$. For nodes x in $\overline{\Omega} \backslash \overline{\Omega}_i^w$ define $\hat{\vartheta}_i^w(x) = 0$. The partition of unity ϑ_i^w is defined as

$$\vartheta_i^w = I_h\left(\frac{\hat{\vartheta}_i^w}{\sum_{j=1}^{N} \hat{\vartheta}_j^w}\right) \quad i = 1, \cdots, N,$$

where I_h is the nodal piecewise linear interpolant on $\mathscr{T}^h(\overline{\Omega})$. Note that the support of ϑ_i^w is Ω_i^{w+1} and $|\nabla \vartheta_i^w| \leq O((w+1)/h)$. We define the weighting diagonal matrix D_i^w as equal to $\vartheta_i^w(x)$ at the nodes x of $\overline{\Omega}$.

3.3 Local Problems

Let us denote by V_i^δ, $i = 1, \cdots, N$, the local space of functions in $H^1(\Omega_i^\delta)$ which are continuous piecewise linear and vanishes only on $\partial\Omega_i^\delta \cap \partial\Omega_D$. For each subdomain Ω_i^δ, let $R_i^\delta : V \to V_i^\delta$ be the regular restriction operator on V_i^δ, that is, $v_i(x) = v(x)$ for nodes $x \in \overline{\Omega}_i^\delta$.

For the local solvers, we respect the original boundary condition and impose Sommerfeld boundary condition on the interior boundaries $\partial\Omega_i^\delta \backslash \partial\Omega$. The associated local projections in matrix form are defined by

$$T_{i,WRAS-H}^\delta = (R_i^\delta D_i^\delta)^T (\tilde{A}_i^\delta)^{-1} R_i^\delta A \quad i = 1, \cdots, N \tag{5}$$

where $\tilde{A}_i^\delta$ are the matrix form of

$$\tilde{a}_i^\delta(u_i, v_i) = \int_{\Omega_i^\delta} (\nabla u_i \cdot \nabla \overline{v}_i - k^2 u_i \overline{v}_i)\, dx + ik \int_{\partial\Omega_i^\delta \backslash (\partial\Omega_D \cup \partial\Omega_N)} u_i \overline{v}_i\, ds. \tag{6}$$

3.4 Coarse Problem

Let c be a nonnegative integer. The coarse space $V_0^{c,p} \in V$ is defined as the space spanned by $D_i^c Q_j^D$ for $i = 1,\ldots,N$ and $j = 1,\cdots,p$. Here, $Q_j := e^{ik\eta_j^T x}$, where $\eta_j = (\cos(\theta_j), \sin(\theta_j))$, with $\theta_j = (j-1)\times\frac{\pi}{p}, j = 1,\cdots,p$, while $Q_j^D(x) := Q_j(x)$ for nodes $x \in \overline{\Omega}\backslash\partial\Omega_D$ and $Q_j^D(x) := 0$ for nodes x on $\partial\Omega_D$. The coarse-to-fine prolongation matrix $(E_0^{c,p})$ consists of columns $D_i^\delta Q_j^D$, while the fine-to-coarse restriction matrix $R_0^{\delta,p}$ consists of rows $(R_i^\delta)^T R_i^\delta Q_j^D$. The first coarse problem we consider in this paper is given by

$$P_{0,RC}^{\delta,c,p} = E_0^{c,p}[R_0^{\delta,p} A E_0^{c,p}]^{-1} R_0^{\delta,p}. \tag{7}$$

3.5 Hybrid Preconditioners

The first preconditioner we consider is given by

$$T_{WRAS-H-RC}^{\delta,c,p} := P_{0,RC}^{\delta,c,p} + (I - P_{0,RC}^{\delta,c,p})(\sum_{i=1}^{N} T_{i,WRAS-H}^{\delta})(I - P_{0,RC}^{\delta,c,p}). \tag{8}$$

Because $P_{0,RC}^{\delta,c,p}$ is a projection, only one coarse problem solver is necessary per iteration of the iterative method.

Other hybrid preconditioners can also be designed. For instance, we can replace the local problem $T_{i,WRAS}^{\delta}$ by

$$P_{i,OBDD-H}^{\delta} := (R_i^\delta D_i^\delta)^T (\tilde{A}_i^\delta)^{-1} R_i^\delta D_i^\delta A$$

or/and replace the coarse problem $P_{0,RC}^{\delta,c,p}$ by something more classical such as

$$P_0^{c,p} = E_0^{c,p}[(E_0^{c,p})^T A E_0^{c,p}]^{-1}(E_0^{c,p})^T.$$

Inserting these operators properly into (7) we obtain preconditioners which we denote by $T_{WRAS-H}^{\delta,c,p}$, $T_{OBDD-H}^{\delta,c,p}$ or $T_{OBDD-H-RC}^{\delta,c,p}$. An interesting structure that $T_{WRAS-H-RC}^{\delta,c,p}$ has, and the others do not, is that the same restriction operators R_i^δ are used to compute the right-hand side for both the local and coarse problems, therefore, computational efficiency can be explored.

4 Shifted Local Operators

The matrix $\tilde{A}_i^\delta$ obtained from the bilinear form (6) can be written as

$$\tilde{A}_i^\delta = A_i^\delta - k^2 M_i^\delta + ikB_i^\delta,$$

where A_i^δ, M_i^δ, and B_i^δ are the corresponding matrices associated to

$$\int_{\Omega_i^\delta} \nabla u_i \cdot \nabla \overline{v}_i \, dx + ik \int_{\partial\Omega_i^\delta \cap \partial\Omega_S} u_i \overline{v}_i \, ds, \quad \int_{\Omega_i^\delta} u_i \overline{v}_i \, dx \quad \text{and} \quad \int_{\partial\Omega_i^\delta \setminus \partial\Omega} u_i \overline{v}_i \, ds,$$

respectively. We note that the local matrix $A_i^\delta - k^2 M_i^\delta$ is singular if k^2 is a generalized eigenvalue of A_i^δ. Alternatively, if we enforce zero Dirichlet boundary condition on the interior boundaries $\partial\Omega_i \cap \Omega_i^\delta$, singularities also might occurs, specially when the subdomains are not small enough. The Sommerfeld term plays the rule of shifting the real spectrum of $A_i^\delta - k^2 M_i^\delta$ to the upper part of the complex plane, therefore, elliminating possible zero eigenvalues. More general shifts were introduced recently by Gijzen et al. [7] and Erlangga et al. [4] to move the spectrum to a disk on the first quadrant. Inspired by this work, we now consider shifts to define the local solvers as

$$\tilde{A}_i^\delta(\alpha_r, \alpha_i, \beta_r, \beta_i) = A_i^\delta + (\alpha_r + i\alpha_i)k^2 M_i^\delta + (\beta_r + i\beta_i) k B_i^\delta, \tag{9}$$

that is, the local Laplacians A_i^δ are shifted by a complex combination of M_i^δ and B_i^δ. Note that $\tilde{A}_i^\delta(-1,0,0,1)$ reduces to the original local solver (6), while $\tilde{A}_i^\delta(-1,0,0,0)$ to $A_i^\delta - k^2 M_i^\delta$.

5 Numerical Results

As a numerical test, we consider a wave guided problem for solving the Helmholtz equation on the unit square. We consider homogeneous Neumann boundary condition on the horizontal sides, homogeneous Sommerfeld on the right vertical side, and a constant identical to one Dirichlet on the left vertical side. The stopping criteria for the PGMRES is to reduce the initial residual by a factor of 10^{-6}. In all tests the right preconditioner is applied.

The triangulation is composed of Courant elements of mesh size $h = 1/256$. The nonoverlapping subdomains Ω_i^0 are squares of size $1/M$, and the number of subdomains is denoted by $nsub = M \times M$. The pair (δ, c) refers to how many layers of elements are used to define the extension of the overlapping subdomains Ω_i^δ and the extension of the support of the coarse basis functions, respectively. The constant k refers to the wave number and p denotes the number of local plane waves used in the coarse space. Table 1 shows that the method $P_{WRAS-H-RC}$ is the most effective method among those introduced in Sect. 3.5. Table 2 shows that we should select the support for the coarse basis functions larger enough, larger than the size of the extended subdomains. Tables 1 and 2 show that the number of iterations decreases when we increase the size of the overlap.

We now test the effectiveness of $P_{WRAS-H-RC}$ for several combinations of local solvers $\tilde{A}_i^\delta(\alpha_r, \alpha_i, \beta_r, \beta_i)$. Table 3 shows results for $\delta = 2$ and Table 4 for $\delta = 0$. We can see from Tables 3 and 4 that the number of iterations using the original local problem are 13 and 34, respectively. It is very surprising and interesting to observe that the number of iterations are 9 and 18 for the combination $(0,1,1,0)$, a

respectable gain in efficiency. Tables 3 and 4 reveal that there exist more effective choices for local solvers rather than the common choice approach of adding a Sommerfeld term on the interior boundary of the subdomains. These preliminary results are very inspiring and encouraging for further numerical and theoretical investigations.

Table 1. The Guided Wave Problem, Sommerfeld boundary condition on interior subdomain boundaries, $n = 257$, $nsub = 64(8 \times 8)$, Tol=10^{-6}, $k = 20$

(δ, c, p)	(0,7,4)	(1,7,4)	(2,7,4)
$OBDD - H$	158	85	43
$WRAS - H$	150	74	36
$OBDD - H - RC$	40	23	16
$WRAS - H - RC$	34	19	13

Table 2. WRAS-H-RC The Guided Wave Problem, Sommerfeld boundary condition on interior subdomain boundaries, $n = 257$, $nsub = 64(8 \times 8)$, $p = 4$, Tol=10^{-6}, $k = 20$

	WRAS-H-RC							
$c=$	1	2	3	4	5	6	7	8
δ =0	78	67	54	46	40	37	34	32
δ =1	190	36	31	25	22	21	19	18
δ =2	181	181	19	18	16	14	13	12

Table 3. The Guided Wave Problem, **WRAS-H-RC** algorithm with Shifted Laplacian local problems, $n = 257$, $nsub = 64$, Tol=10^{-6}, $p = 4$, $k = 20$, $c = 7$, $\delta = 2$

	$\alpha_r =$	-1	-1	-1	0	0	0	1	1	1
	$\alpha_i =$	-1	0	1	-1	0	1	-1	0	1
$\beta_r = -1$	$\beta_i = -1$	37	53	116	22	28	210	17	22	48
$\beta_r = -1$	$\beta_i = 0$	236	123	199	154	275	139	105	300*	138
$\beta_r = -1$	$\beta_i = 1$	66	34	28	227	24	16	55	22	17
$\beta_r = 0$	$\beta_i = -1$	20	23	62	14	14	20	12	11	12
$\beta_r = 0$	$\beta_i = 0$	19	16	13	17	300*	12	14	13	10
$\beta_r = 0$	$\beta_i = 1$	55	**13**	13	23	13	11	15	12	11
$\beta_r = 1$	$\beta_i = -1$	15	12	12	13	10	10	12	10	9
$\beta_r = 1$	$\beta_i = 0$	13	17	11	12	10	9	12	10	**8**
$\beta_r = 1$	$\beta_i = 1$	17	10	11	12	10	9	11	10	9

Table 4. The Guided Wave Problem, **WRAS-H-RC** algorithm with Shifted Laplacian local problems, $n = 257$, $nsub = 64$, Tol=10^{-6}, $p = 4$, $k = 20$, $c = 7$, $\delta = 0$

	$\alpha_r =$	-1	-1	-1	0	0	0	1	1	1
	$\alpha_i =$	-1	0	1	-1	0	1	-1	0	1
$\beta_r = -1$	$\beta_i = -1$	168	213	300*	99	168	300*	69	106	300*
$\beta_r = -1$	$\beta_i = 0$	291	207	243	238	300*	209	221	300*	300*
$\beta_r = -1$	$\beta_i = 1$	300*	137	101	300*	130	63	300*	107	67
$\beta_r = 0$	$\beta_i = -1$	55	69	289	38	42	80	34	30	32
$\beta_r = 0$	$\beta_i = 0$	45	31	30	38	300*	27	34	24	24
$\beta_r = 0$	$\beta_i = 1$	279	**34**	33	94	39	30	40	35	31
$\beta_r = 1$	$\beta_i = -1$	34	31	39	29	25	22	27	24	21
$\beta_r = 1$	$\beta_i = 0$	27	22	21	**24**	20	**18**	24	21	20
$\beta_r = 1$	$\beta_i = 1$	51	23	21	25	21	20	23	21	21

Bibliography

[1] Xiao-Chuan Cai and Marcus Sarkis. A restricted additive Schwarz preconditioner for general sparse linear systems. *SIAM Journal on Scientific Computing*, 21:239–247, 1999.

[2] Xiao-Chuan Cai, Mario A. Casarin, Frank W. Elliott Jr., and Olof B. Widlund. Overlapping Schwarz algorithms for solving Helmholtz's equation. In *Domain decomposition methods, 10 (Boulder, CO, 1997)*, pages 391–399. Amer. Math. Soc., Providence, RI, 1998.

[3] B. Deprés. *Méthodes de décomposition de domains pour les problèms de propagation d'ondes en régime harmonique*. PhD thesis, Université Paris IX Dauphine, 1991.

[4] Y.A. Erlangga, C. Vuik, and C.W. Oosterlee. On a class of preconditioners for solving the Helmholtz equation. *Applied Numerical Mathematics*, 50:409–425, 2004.

[5] C. Farhat, A. Macedo, and M. Lesoinne. A two-level domain decomposition method for the iterative solution of high-frequency exterior Helmholtz problems. *Numer. Math.*, 85(2):283–303, 2000.

[6] Charbel Farhat, Jan Mandel, and Francois-Xavier Roux. Optimal convergence properties of FETI domain decomposition method. *Comput. Methods Appl. Mech. Eng.*, 115:367–388, 1994.

[7] M.B. Van Gijzen, Y.A. Erlangga, and C. Vuik. Spectral analysis of the discrete Helmholtz operato preconitioned with a shifted Laplacian. *SIAM J. Sci. Comput.*, 29(5):1942–1958, 2007.

[8] Jung-Han Kimn. A convergence theory for an overlapping Schwarz algorithm using discontinuous iterates. *Numer. Math.*, 100(1):117–139, 2005.

[9] Jung-Han Kimn and Marcus Sarkis. OBDD: Overlapping balancing domain decomposition methods and generalizations to the helmholtz equation. In David

Keyes and Olof B. Widlund, editors, *Domain Decomposition Methods in Science and Engineering XVI*, volume 55 of *Lecture Notes in Computational Science and Engineering*, pages 317–324. Springer-Verlag, 2006.

[10] Jung-Han Kimn and Marcus Sarkis. Restricted overlapping balancing domain decomposition methods and restricted coarse probelms for the Helmholtz equation. *Comput. Methods Appl. Mech. Engrg.*, 106:1507–1514, 2007.

[11] Jan Mandel. Balancing domain decomposition. *Comm. Numer. Meth. Engrg.*, 9:233–241, 1993.

[12] Jens Markus Melenk. *On Generalized Finite Element Methods*. PhD thesis, The University of Maryland, 1995.

[13] Marcus Sarkis. Partition of unity coarse spaces: Discontinuous coefficients, multi-level versions and applications to elasticity. In Ismael Herrera, David E. Keyes, Olof B. Widlund, and Robert Yates, editors, *14th International Conference on Domain Decomposition Methods, Cocoyoc, Mexico*, 2002.

A Subspace Correction Method for Nearly Singular Linear Elasticity Problems

E. Karer[1], J. K. Kraus[1], and L. T. Zikatanov[2]

[1] Johann Radon Institute for Computational and Applied Mathematics, Austrian Academy of Sciences, Altenberger Strasse 69, 4040 Linz, Austria {erwin.karer, johannes.kraus}@oeaw.ac.at

[2] Penn State University, State College ludmil@psu.edu

1 Introduction

The focus of this work is on constructing a robust (uniform in the problem parameters) iterative solution method for the system of linear algebraic equations arising from a nonconforming finite element discretization based on reduced integration. We introduce a specific space decomposition into two overlapping subspaces that serves as a basis for devising a uniformly convergent subspace correction algorithm. We consider the equations of linear elasticity in primal variables. For nearly incompressible materials, i.e., when the Poisson ratio ν approaches $1/2$, this problem becomes ill-posed and the resulting discrete problem is nearly singular.

Subspace correction methods for nearly singular systems have been studied in [10] leading to robust multigrid methods for planar linear elasticity problems (see [11]). In [13] a multigrid method has been presented for a finite element discretization with $P_2 - P_0$ elements. This approach relies on a local basis for the weakly divergence-free functions.

In this setting, presently known (multilevel) iterative solution methods are optimal or nearly optimal for the pure displacement problem only, i.e., when Dirichlet boundary conditions are imposed on the entire boundary, see, e.g., [1, 4]. For pure traction or mixed boundary conditions the problem gets more involved. It is known, that standard (conforming and nonconforming) finite element methods then require certain stabilization techniques, see, e.g., [3, 6]. We employ a discretization scheme introduced in [3] which achieves the stabilization via reduced integration. Note that based on an appropriate discrete version of Korn's second inequality optimal error estimates have been shown for this method (see [3]).

The remainder of this paper is organized as follows: The formulation of the linear elasticity problem with pure traction boundary conditions and its finite element discretization are given in Sect. 2. We briefly recall some convergence results for the *Method of Successive Subspace Correction* (MSSC) in Sect. 3. In Sect. 4 we present a specific space decomposition which defines an MSSC preconditioner. Finally, we

R. Bank et al. (eds.), *Domain Decomposition Methods in Science and Engineering XX*, Lecture Notes in Computational Science and Engineering 91,
DOI 10.1007/978-3-642-35275-1_17,

present a numerical test illustrating the optimal performance of the preconditioner in Sect. 5.

2 Problem Formulation

For the sake of simplicity we consider only two-dimensional problems in this paper. Let Ω be a bounded, connected and open subset of $\mathbb{R}^2$, denoting the reference configuration of an elastic body. The boundary of Ω is denoted by $\partial\Omega$. Following [3] we consider the pure traction problem of linear elasticity which reads

$$\boldsymbol{\sigma} = \mu\left[\boldsymbol{\varepsilon}(\mathbf{u}) + \frac{\nu}{1-2\nu}\operatorname{div}\mathbf{u} I\right] \qquad \text{in } \Omega, \tag{1a}$$

$$-\operatorname{div}\boldsymbol{\sigma} = \mathbf{f} \qquad \text{in } \Omega, \tag{1b}$$

$$\boldsymbol{\sigma}\cdot\mathbf{n} = \mathbf{g} \qquad \text{on } \partial\Omega. \tag{1c}$$

where $\boldsymbol{\sigma}$ denotes the stress tensor and $\boldsymbol{\varepsilon}(\mathbf{u}) := \nabla^{(s)}\mathbf{u}$ is the symmetric gradient, i.e., $\varepsilon_{ij}(\mathbf{u}) := \frac{1}{2}\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right)$. Further $\mathbf{u}$ denotes the vector of displacements, $\mathbf{f}$ denotes the body forces, $\mathbf{n}$ is the outwards pointing unit normal vector on $\Gamma = \partial\Omega$ and $\mathbf{g}$ is the applied load on Γ. The properties of the material depend on the Poisson ratio $\nu \in [0, 1/2)$, and the shear modulus $\mu := \frac{E}{1+\nu}$ where E is the modulus of elasticity.

We consider the space $\boldsymbol{V}^{\mathrm{RBM}} := \{\mathbf{v} : \mathbf{v} = (a_1 + by, a_2 - bx)^t,\ a_1, a_2, b \in \mathbb{R}\}$ of rigid body motions and define the subspace $\hat{\boldsymbol{V}}$ of H^1-functions orthogonal to $\boldsymbol{V}^{\mathrm{RBM}}$, i.e.,

$$\hat{\boldsymbol{V}} := \{\mathbf{v} \in [H^1(\Omega)]^2 : \int_\Omega \mathbf{v}\,\mathrm{d}\mathbf{x} = \mathbf{0} \quad \text{and} \quad \int_\Omega v_1 y - v_2 x\,\mathrm{d}\mathbf{x} = 0\}. \tag{2}$$

Let $\mathscr{T}_H$ be a quasi-uniform triangulation of Ω. Moreover, we subdivide each triangle $T \in \mathscr{T}_H$ into four congruent triangles by adding the midpoints of the edges to the set of vertices. The obtained refined triangulation $\mathscr{T}_h$ of Ω has a mesh size $h = H/2$. We introduce the vector space $\boldsymbol{V} := [V]^2 := [H^1(\Omega)]^2$ and the subspace $\boldsymbol{V}_h := [V_h]^2$, which consists of the vector-valued continuous piecewise linear functions on the fine mesh $\mathscr{T}_h$. Next we define $\hat{\boldsymbol{V}}_h := \boldsymbol{V}_h \cap \hat{\boldsymbol{V}}$ and denote the space of piecewise constant functions on $\mathscr{T}_H$ by S_H. Then we consider the problem: Find $\mathbf{u}_h \in \hat{\boldsymbol{V}}_h$ such that

$$a(\mathbf{u}_h, \mathbf{v}_h) = L(\mathbf{v}_h) := (\mathbf{f}, \mathbf{v}_h)_0 + \int_{\partial\Omega} \mathbf{g}\cdot\mathbf{v}_h\,\mathrm{d}s \qquad \forall \mathbf{v}_h \in \hat{\mathbf{V}}_h, \tag{3}$$

$$a(\mathbf{u}_h, \mathbf{v}_h) := \mu\left((\boldsymbol{\varepsilon}(\mathbf{u}_h), \boldsymbol{\varepsilon}(\mathbf{v}_h))_0 + \frac{\nu}{1-2\nu}(P_0\operatorname{div}\mathbf{u}_h, P_0\operatorname{div}\mathbf{v}_h)_0\right), \tag{4}$$

where $\mathbf{f} \in [L_2(\Omega)]^2$ and $\mathbf{g} \in [L_2(\partial\Omega)]^2$. P_0 is the L^2-projection onto S_H, that is,

$$P_0(v)|_{T_H} = \frac{1}{|T_H|}\int_{T_H} v\,\mathrm{d}\mathbf{x} \qquad \forall T_H \in \mathscr{T}_H, \tag{5}$$

for any scalar function $v \in L^2(\Omega)$. It is known that under the compatibility condition $L(\mathbf{v}) = 0$ for all $\mathbf{v} \in \boldsymbol{V}^{\mathrm{RBM}}$ problem (3) has a unique solution $\mathbf{u}_h \in \hat{\boldsymbol{V}}_h$, see, e.g., [1]. In [3] optimal order error estimates have been shown for this approximation, which are robust with respect to the Poisson ratio ν.

3 Subspace Correction Framework

The general framework of subspace correction methods is closely related to the abstract Schwarz theory, see, e.g., [5, 14].

Let us consider the variational problem: Find $u \in V$ such that

$$a(u, v) = f(v) \qquad \forall v \in V\,, \tag{6}$$

with $V \subset H$ being a closed subset of the Hilbert space H. Moreover, we assume that the bilinear form $a(.,.) : H \times H \to \mathbb{R}$ is continuous, symmetric, and H-elliptic. If f is a continuous linear functional on H, then this problem is well-posed.

Now, let us split V into a–not necessarily direct–sum of closed subspaces $V_i \subset V$, $i = 1, \ldots, J$, i.e., $V = \sum_{i=1}^{J} V_i$. With each subspace V_i we associate a symmetric, bounded, and elliptic bilinear form $a_i(.,.)$ approximating $a(.,.)$ on V_i. The MSSC (see [16, Algorigthm 2.1]) solves the residual equation for $i = 1, \ldots, J$ with $u_l = u^l$: Find $e_i \in V_i$ such that for all $v_i \in V_i$, there holds:

$$a(e_i, v_i) = f(v_i) - a(u_{l+i-1}, v_i), \quad \text{and set} \quad u_{l+i} = u_{l+i-1} + e_i, \tag{7}$$

Finally, the next iterate is $u^{l+1} = u_{l+J}$. Let $T_i : V \to V_i$ be defined as

$$a_i(T_i v, v_i) = a(v, v_i), \quad \text{for all} \quad v_i \in V_i.$$

The assumptions on $a_i(.,.)$ imply that T_i is well-defined, $\mathscr{R}(T_i) = V_i$, and $T_i : V_i \to V_i$ is an isomorphism. The error after l iterations of the MSSC is given by $u - u^l = E(u - u^{l-1}) = \ldots = E^l(u - u^0)$, where the error propagation operator E can be represented in product form , i.e.,

$$E = (I - T_J)(I - T_{J-1}) \cdots (I - T_1)\,. \tag{8}$$

In the following we consider the case of exact subspace solves, i.e., $a_i(.,.) = a(.,.)$ on V_i, in which T_i reduces to the idempotent, a-adjoint operator P_i defined by

$$a(P_i v, v_i) = a(v, v_i) \qquad \forall v_i \in V_i\,. \tag{9}$$

For a proof of the following identity for the energy norm of the error propagation operator we refer the reader to [16].

Theorem 1. *Under the assumptions* (9) *and* $V = \sum_{i=1}^{J} V_i$ *we have*

$$\|E\|_a^2 = \|(I - P_J)(I - P_{J-1}) \cdots (I - P_1)\|_a^2 = \frac{c_0}{1 + c_0} \tag{10}$$

where $c_0 = \sup_{\|v\|_a = 1} \inf_{\sum_i v_i = v} \sum_{i=1}^{J} \|P_i \sum_{j=i+1}^{J} v_j\|_a^2 < \infty$.

Let $\mathscr{E}_H$ be the set of edges of $\mathscr{T}_H$ and $\mathscr{V}_H$ be the set of (coarse) vertices of the mesh $\mathscr{T}_H$. Then for any vertex $v_i \in \mathscr{V}_H$ we denote the set of edges sharing v_i by $\mathscr{N}_i^{\mathscr{E}}$. For any edge $E = (v_{E,1}, v_{E,2}) \in \mathscr{E}_H$ by φ_E we denote the scalar nodal basis function corresponding to the midpoint of the edge E, and by $\varphi_{E,1}$ and $\varphi_{E,2}$ the nodal basis

functions corresponding to the vertices $v_{E,1}$ and $v_{E,2}$ of E. The corresponding vector-valued degrees of freedom (dof) of any function $\mathbf{v}_h \in \boldsymbol{V}_h$ are denoted by $\mathbf{v}_E$, $\mathbf{v}_{E,1}$ and $\mathbf{v}_{E,2}$, respectively. We further use φ_i and $\mathbf{v}_i$ to denote the basis functions and dof associated with the vertices from $\mathscr{V}_H$.

For any edge $E \in \mathscr{E}_H$ we assume that $v_{E,1} < v_{E,2}$ and that the globally defined tangential vector τ_E points from $v_{E,1}$ to $v_{E,2}$. The global edge normal vector $\mathbf{n}_E$ is orthogonal to τ_E and is obtained from τ_E by a clockwise rotation. By $\boldsymbol{V}_H^{RT}$ we denote the lowest order Raviart Thomas space (cf. [2]), i.e.,

$$\boldsymbol{V}_H^{RT} := \{\mathbf{v} \in [L^2(\Omega)]^2 : \mathbf{v} = \mathbf{a} + (bx, by)^t \text{ on each } T \in \mathscr{T}_H,\ \mathbf{a} \in \mathbb{R}^2,\ b \in \mathbb{R}\} \quad (11)$$

where the degrees of freedom are the normal fluxes over the edges E, i.e., $F_E^{RT}(\mathbf{v}) := \frac{1}{|E|}\int_E \mathbf{v}\cdot\mathbf{n}_E\,\mathrm{d}s$. The basis functions $\boldsymbol{\varphi}_E^{RT}$ corresponding to an edge E of an element $T \in \mathscr{T}_H$ are such that $F_{E'}^{RT}(\boldsymbol{\varphi}_E^{RT}) := \delta_{EE'}$. We also use the projection $\Pi^{RT} : \boldsymbol{V} \mapsto \boldsymbol{V}_H^{RT}$ defined by $\Pi^{RT}(\mathbf{v}) = \sum_{E\in\mathscr{E}_H} F_E^{RT}(\mathbf{v})\boldsymbol{\varphi}_E^{RT}$, for which the commuting property $P_0 \operatorname{div}\mathbf{v}_h = \operatorname{div}\Pi^{RT}(\mathbf{v}_h)$ holds for any $\mathbf{v}_h \in \boldsymbol{V}_h$ (cf. [2, p. 131]).

4 Space Decomposition

Let us consider the following unique decomposition of any function $\mathbf{v}_h \in \boldsymbol{V}_h$:

$$\begin{aligned}
\mathbf{v}_h &= \sum_{i\in\mathscr{V}_H} \varphi_i \mathbf{v}_i + \sum_{E\in\mathscr{E}_H} \varphi_E \mathbf{v}_E \\
&= \underbrace{\sum_{i\in\mathscr{V}_H} \left[\varphi_i \mathbf{v}_i - \frac{1}{2}\sum_{E\in\mathscr{N}_i^{\mathscr{E}}} (\mathbf{v}_i\cdot\mathbf{n}_E)\varphi_E\mathbf{n}_E\right]}_{=:\mathbf{v}_\mathscr{V}} + \underbrace{\sum_{E\in\mathscr{E}_H} (\mathbf{v}_E\cdot\tau_E)\varphi_E\tau_E}_{=:\mathbf{v}_\tau} \\
&\quad + \underbrace{\sum_{E\in\mathscr{E}_H} \left(\left[\mathbf{v}_E + \frac{1}{2}(\mathbf{v}_{E,1}+\mathbf{v}_{E,2})\right]\cdot\mathbf{n}_E\right)\varphi_E\mathbf{n}_E}_{=:\mathbf{v}_n}.
\end{aligned}$$

Next we define the splitting $\boldsymbol{V}_h = \boldsymbol{V}_\mathscr{V} \oplus \boldsymbol{V}_\tau \oplus \boldsymbol{V}_n$, where

$$\boldsymbol{V}_\mathscr{V} := \{\mathbf{v}_h \in \boldsymbol{V}_h : \mathbf{v}_h = \sum_{i\in\mathscr{V}_H}\left[\varphi_i\mathbf{v}_i - \frac{1}{2}\sum_{E\in\mathscr{N}_i^{\mathscr{E}}}(\mathbf{v}_i\cdot\mathbf{n}_E)\,\varphi_E\mathbf{n}_E\right]\},$$

$$\boldsymbol{V}_\tau := \{\mathbf{v}_h \in \boldsymbol{V}_h : \mathbf{v}_h = \sum_{E\in\mathscr{E}_H}\alpha_E\varphi_E\tau_E\},\quad \boldsymbol{V}_n := \{\mathbf{v}_h \in \boldsymbol{V}_h : \mathbf{v}_h = \sum_{E\in\mathscr{E}_H}\alpha_E\varphi_E\mathbf{n}_E\}.$$

Note that $\Pi^{RT}(\boldsymbol{V}_\mathscr{V}) = \Pi^{RT}(\boldsymbol{V}_\tau) = \{0\}$. Next, we introduce the spaces

$$\boldsymbol{V}_{\mathrm{curl}} := \{\mathbf{v}_h \in \boldsymbol{V}_h : \mathbf{v}_h = \sum_{i\in\mathscr{V}_H}\beta_i\sum_{E\in\mathscr{N}_i^{\mathscr{E}}}\frac{\delta_{E,i}}{|E|}\varphi_E\mathbf{n}_E\},$$

$$\boldsymbol{V}_{\nabla_h} := \{\mathbf{v}_h \in \boldsymbol{V}_h : \mathbf{v}_h = \sum_{T\in\mathscr{T}_H}\gamma_T\sum_{E\subset T}(\mathbf{n}_E\cdot\mathbf{n}_{E,T})\,\varphi_E\mathbf{n}_E\}.$$

Here $\delta_{E,i}$ is defined by

$$\delta_{E,i} = \begin{cases} -1 \text{ if } i = v_{E,1} \\ 1 \text{ if } i = v_{E,2} \end{cases} . \tag{12}$$

Note that $\boldsymbol{V}_{\text{curl}} \subset \boldsymbol{V}_n$, and $\boldsymbol{V}_{\nabla_h} \subset \boldsymbol{V}_n$, and the following properties hold:

$$\begin{aligned} P_0 \operatorname{div}(\mathbf{v}_{\text{curl}}) = \operatorname{div} \Pi^{RT}(\mathbf{v}_{\text{curl}}) = 0 \qquad & \forall \mathbf{v}_{\text{curl}} \in \boldsymbol{V}_{\text{curl}}, \\ P_0 \operatorname{div}(\mathbf{v}_{\nabla_h}) = \operatorname{div} \Pi^{RT}(\mathbf{v}_{\nabla_h}) \neq 0 \qquad & \forall \mathbf{v}_{\nabla_h} \in \boldsymbol{V}_{\nabla_h}. \end{aligned}$$

Moreover, $\dim(\boldsymbol{V}_{\text{curl}}) = n_{v,H} - 1$ and $\dim(\boldsymbol{V}_{\nabla_h}) = n_{T,H}$, and thus, using Euler's formula, i.e., $n_{v,H} - 1 + n_{T,H} = n_{E,H}$, we find that $\boldsymbol{V}_n = \boldsymbol{V}_{\text{curl}} \oplus \boldsymbol{V}_{\nabla_h}$. Hence we obtain

$$\boldsymbol{V}_h = \boldsymbol{V}_{\mathscr{V}} \oplus \boldsymbol{V}_{\tau} \oplus \boldsymbol{V}_{\text{curl}} \oplus \boldsymbol{V}_{\nabla_h}. \tag{13}$$

Finally, we decompose $\boldsymbol{V}_h$ into two overlapping subspaces $\boldsymbol{V}_I$ and $\boldsymbol{V}_{II}$:

$$\boldsymbol{V}_I = \boldsymbol{V}_{\mathscr{V}} \oplus \boldsymbol{V}_{\tau} \oplus \boldsymbol{V}_{\text{curl}} \tag{14}$$

$$\boldsymbol{V}_{II} = \boldsymbol{V}_{\tau} \oplus \boldsymbol{V}_{\text{curl}} \oplus \boldsymbol{V}_{\nabla_h} \tag{15}$$

The overlap of $\boldsymbol{V}_I$ and $\boldsymbol{V}_{II}$ is given by $\boldsymbol{V}_{\tau} + \boldsymbol{V}_{\text{curl}}$, and any element $\mathbf{v}_{II} \in \boldsymbol{V}_{II}$ can be uniquely decomposed into $\mathbf{v}_{II} = \mathbf{v}_{\tau} + \mathbf{v}_{\text{curl}} + \mathbf{v}_{\nabla_h}$, with $\mathbf{v}_{\tau} \in \boldsymbol{V}_{\tau}$, $\mathbf{v}_{\text{curl}} \in \boldsymbol{V}_{\text{curl}}$ and $\mathbf{v}_{\nabla_h} \in \boldsymbol{V}_{\nabla_h}$. However, finding the components $\mathbf{v}_{\text{curl}} \in \boldsymbol{V}_{\text{curl}}$ and $\mathbf{v}_{\nabla_h} \in \boldsymbol{V}_{\nabla_h}$ for a given function $\mathbf{v}_n \in \boldsymbol{V}_n$ requires a solution of a system with an M-matrix corresponding to the lowest order mixed method for Laplace equation with lumped mass [2].

Note that since $P_0 \operatorname{div}(\boldsymbol{V}_I) = \operatorname{div} \Pi^{RT}(\boldsymbol{V}_I) = \{0\}$ the bilinear form $a(.,.)$ satisfies

$$a(\mathbf{u}_I, \mathbf{v}_I) = \mu(\boldsymbol{\varepsilon}(\mathbf{u}_I), \boldsymbol{\varepsilon}(\mathbf{v}_I))_0 \qquad \forall \mathbf{u}_I, \mathbf{v}_I \in \boldsymbol{V}_I, \tag{16}$$

and in the limit case $\nu = 0$ we have $a(\mathbf{u}_h, \mathbf{v}_h) = \mu(\boldsymbol{\varepsilon}(\mathbf{u}_h), \boldsymbol{\varepsilon}(\mathbf{v}_h))_0$ for all $\mathbf{u}_h, \mathbf{v}_h \in \boldsymbol{V}_h$.

In the following, we use the operator representations $A : V \to V$ and $A_{\varepsilon} : V \to V$ for the bilinear forms $a(.,.)$ and $\mu(\boldsymbol{\varepsilon}(.), \boldsymbol{\varepsilon}(.))_0$. If we symmetrize the MSSC, we obtain the following error propagation $\bar{E}_{MSSC}$, compare with (8) in case of $J = 2$ and exact subsolves, i.e.,

$$\bar{E}_{MSSC} = (I - P_I)(I - P_{II})(I - P_I).$$

The error propagation operator can be rewritten as $\bar{E}_{MSSC} = I - \bar{B}_{MSSC} A$, with symmetric $\bar{B}_{MSSC}$. Further, $\bar{B}_{MSSC}$ is positive definite, since $\bar{E}_{MSSC}$ is non-expansive. Note that even though $\bar{B}_{MSSC} = (I - \bar{E}_{MSSC}) A^{-1}$ formally involves the inverse of A, we do not need A^{-1} in order to apply $\bar{B}_{MSSC}$.

If ν is bounded away from the incompressible limit $1/2$, we know that A_{ε} is spectrally equivalent to A. Further, there are efficient preconditioners for A_{ε}. We now define the additive preconditioner B by

$$B := \frac{1 - 2\nu}{1 - \nu} A_{\varepsilon}^{-1} + \frac{\nu}{1 - \nu} \bar{B}_{MSSC}. \tag{17}$$

Note that B is a convex combination of A_{ε}^{-1} and $\bar{B}_{MSSC}$.

Remark 1. It has been shown in [14, 16] that an inexact solution of the subproblems (7) results in a uniform preconditioner under reasonable assumptions. The subproblems on the spaces $\boldsymbol{V}_I$ and $\boldsymbol{V}_h$ involve the bilinear form

$$\bar{a}(\mathbf{u}_i, \mathbf{v}_i) = \mu(\boldsymbol{\varepsilon}(\mathbf{u}_i), \boldsymbol{\varepsilon}(\mathbf{v}_i))_0 \qquad \forall \mathbf{u}_i, \mathbf{v}_i \in \boldsymbol{W} = \boldsymbol{V}_I, \boldsymbol{V}_h. \tag{18}$$

Any efficient preconditioning technique for the vector-Laplace equation can be employed in these steps, e.g., classical AMG (see [12]) or AMGm (see [8]).

The problem on $\boldsymbol{V}_{II} = \boldsymbol{V}_E := \{\mathbf{v}_h \in \boldsymbol{V}_h : \mathbf{v}_h(\mathbf{x}_i) = \mathbf{0}\ v_i \in \mathcal{V}_H\}$ is more involved. First, by using Korn's inequality, Poincarè's inequality and the inverse inequality one can show that

$$\|\boldsymbol{\varepsilon}(\mathbf{v}_E)\|_0^2 \approx \|\boldsymbol{\nabla}\mathbf{v}_E\|_0^2 \approx H^{-2}\|\mathbf{v}_E\|_0^2.$$

Second, any function $\mathbf{v}_E \in \boldsymbol{V}_E$ can be uniquely decomposed into $\mathbf{v}_E = \mathbf{v}_n + \mathbf{v}_\tau$ where $\mathbf{v}_n \in \boldsymbol{V}_n$ and $\mathbf{v}_\tau \in \boldsymbol{V}_\tau$. Moreover, by locally estimating the angle between $\boldsymbol{V}_n$ and $\boldsymbol{V}_\tau$ in the $a(\cdot,\cdot)$-inner product, it can be shown that

$$\|\mathbf{v}_E\|_0^2 = \|\mathbf{v}_n + \mathbf{v}_\tau\|_0^2 \approx \|\mathbf{v}_n\|_0^2 + \|\mathbf{v}_\tau\|_0^2 \tag{19}$$

holds uniformly with respect to the mesh size h. Furthermore $\Pi^{RT}(\mathbf{v}_\tau) = 0$ for all $\mathbf{v}_\tau \in \boldsymbol{V}_\tau$. Hence, the relation $a(\mathbf{u}_E, \mathbf{v}_E) \approx \tilde{a}(\mathbf{u}_E, \mathbf{v}_E)$ holds on $\boldsymbol{V}_{II}$ where

$$\begin{aligned}\tilde{a}(\mathbf{u}_E, \mathbf{v}_E) := \mu\Big\{ & H^{-2}(\mathbf{u}_\tau, \mathbf{v}_\tau)_0 \\ & + H^{-2}(\mathbf{u}_n, \mathbf{v}_n)_0 + \frac{\nu}{1-2\nu}(P_0 \operatorname{div}\mathbf{u}_n, P_0 \operatorname{div}\mathbf{v}_n)_0\Big\}.\end{aligned} \tag{20}$$

Now, using the interpolation operator $I_{RT}^h : \boldsymbol{V}_H^{RT} \to \boldsymbol{V}_h$, defined by $I_{RT}^h(\varphi_E^{RT}) = 2\varphi_E \mathbf{n}_E \in \boldsymbol{V}_n$, one can show that $\boldsymbol{V}_n$ is isomorphic to $\boldsymbol{V}_H^{RT}$. Thus solving a variational problem with $\tilde{a}(.,.)$ on $\boldsymbol{V}_n$ is equivalent to solving a problem with the bilinear form

$$a_{RT}(\mathbf{u}_{RT}, \mathbf{v}_{RT}) := \mu\left\{H^{-2}(\mathbf{u}_{RT}, \mathbf{v}_{RT})_0 + \frac{\nu}{1-2\nu}(\operatorname{div}\mathbf{u}_{RT}, \operatorname{div}\mathbf{v}_{RT})_0\right\}, \tag{21}$$

on $\boldsymbol{V}_H^{RT}$ (see [7, 15]). An efficient solver for the latter problem can be designed by using the auxiliary space preconditioner of [7], or by using the robust algebraic multilevel iteration method developed in [9].

5 Numerical Experiment

We now perform a numerical test to show that the preconditioner (17) is an efficient and robust preconditioner. We consider the problem with homogenous Dirichlet boundary conditions on the unit square $\Omega = (0,1)^2$. The number of PCG iterations for a residual reduction by a factor 10^8 are shown in Table 1. The subproblems on V_I and V_{II} are solved exactly. Additionally, we list the estimated condition numbers $\kappa(BA)$, obtained from the Lanczos process.

Table 1. Iteration numbers (#it.) and condition numbers ($\kappa(BA)$) of the pcg-cycle.

#DOF	242		1058		4418		18050		72962		293378	
	#it.	κ	#it.	κ	#it.	κ	#it.	κ	#it.	κ	#it.	κ
$\nu = 0$:	1	1.00	1	1.00	1	1.00	1	1.00	1	1.00	1	1.00
$\nu = 0.25$:	8	1.41	8	1.48	8	1.53	9	1.55	9	1.57	9	1.57
$\nu = 0.4$:	10	1.90	11	2.19	12	2.38	12	2.49	13	2.57	13	2.62
$\nu = 0.45$:	11	2.11	12	2.61	14	3.01	15	3.25	15	3.41	15	3.52
$\nu = 0.49$:	10	1.90	11	2.54	14	3.31	16	3.97	17	4.39	17	4.69
$\nu = 0.499$:	9	1.98	10	1.98	11	2.13	14	2.99	15	3.83	17	4.51
$\nu = 0.4999$:	9	1.99	9	1.99	9	1.99	10	1.99	12	2.43	13	3.34
$\nu = 0.49999$:	9	1.99	9	1.99	9	2.00	9	2.00	9	2.00	10	2.00

Acknowledgments The authors gratefully acknowledge the support by the Austrian Academy of Sciences and by the Austrian Science Fund (FWF), Project No. P19170-N18 and by the National Science Foundation NSF-DMS 0810982.

Bibliography

[1] S.C. Brenner and R.L. Scott. *The Mathematical Theory of Finite Element Methods (Texts in Applied Mathematics)*. Springer, 3rd edition, December 2007.

[2] F. Brezzi and M. Fortin. *Mixed and Hybrid Finite Element Methods*. Springer-Verlag New York Inc., 1991.

[3] R.S. Falk. Nonconforming finite element methods for the equations of linear elasticity. *Math. Comp.*, 57(196):529–550, 1991.

[4] I. Georgiev, J.K. Kraus, and S. Margenov. Multilevel preconditioning of Crouzeix-Raviart 3D pure displacement elasticity problems. In I. Lirkov et al., editors, *LSSC*, volume 5910 of *LNCS*, pages 103–110. Springer, 2010.

[5] M. Griebel and P. Oswald. On the abstract theory of additive and multiplicative Schwarz algorithms. *Numer. Math.*, 70(2):163–180, 1995.

[6] P. Hansbo and M.G. Larson. Discontinuous Galerkin and the Crouzeix-Raviart element: application to elasticity. *Math. Model. Numer. Anal.*, 37(1):63–72, 2003.

[7] R. Hiptmair and J. Xu. Nodal auxiliary space preconditioning in H(curl) and H(div) spaces. *SIAM J. Numer. Anal.*, 45(6):2483–2509, 2007.

[8] E. Karer and J. K. Kraus. Algebraic multigrid for finite element elasticity equations: Determination of nodal dependence via edge-matrices and two-level convergence. *Int. J. Numer. Meth. Engng.*, 83(5):642–670, 2010.

[9] J.K. Kraus and S.K. Tomar. Algebraic multilevel iteration method for lowest order Raviart-Thomas space and applications. *Int. J. Numer. Meth. Engng.*, 86 (10):1175–1196, 2011.

[10] Y.-J. Lee, J. Wu, J. Xu, and L. T. Zikatanov. Robust subspace correction methods for nearly singular systems. *Math. Models Methods Appl. Sci.*, 17(11): 1937–1963, 2007.

[11] Y.-J. Lee, J. Wu, and J. Chen. Robust multigrid method for the planar linear elasticity problems. *Numerische Mathematik*, 113:473–496, 2009.

[12] J.W. Ruge and K. Stüben. Algebraic multigrid (AMG). In S. F. McCormick, editor, *Multigrid Methods*, volume 3 of *Frontiers Appl. Math.*, pages 73–130, Philadelphia, 1987. SIAM.

[13] J. Schöberl. Multigrid methods for a parameter dependent problem in primal variables. *Numer. Math.*, 84(1):97–119, 1999.

[14] J. Xu. Iterative methods by space decomposition and subspace correction. *SIAM Review*, 34(4):581–613, 1992.

[15] J. Xu. The auxiliary space method and optimal multigrid preconditioning techniques for unstructured grids. *Computing*, 56:215–235, 1996.

[16] J. Xu and L. Zikatanov. The method of alternating projections and the method of subspace corrections in Hilbert space. *J. Amer. Math. Soc.*, 15(3):573–597, 2002.

Adaptive Finite Element Methods with Inexact Solvers for the Nonlinear Poisson-Boltzmann Equation

Michael Holst[1], Ryan Szypowski[2], and Yunrong Zhu[3]

[1] Departments of Mathematics and Physics, University of California San Diego, La Jolla, CA 92093. Supported in part by NSF Awards 0715146 and 0915220, and by CTBP and NBCR, mholst@math.ucsd.edu, http://ccom.ucsd.edu/~mholst/

[2] Department of Mathematics and Statistics, California State Polytechnic University, Pomona, Pomona, CA 91768. Supported in part by NSF Award 0715146, rsszypowski@csupomona.edu

[3] Department of Mathematics, Idaho State University, Pocatello, 83209-8085, zhuyunr@isu.edu

1 Introduction

In this article we study adaptive finite element methods (AFEM) with inexact solvers for a class of semilinear elliptic interface problems. We are particularly interested in nonlinear problems with discontinuous diffusion coefficients, such as the nonlinear Poisson-Boltzmann equation and its regularizations. The algorithm we study consists of the standard SOLVE-ESTIMATE-MARK-REFINE procedure common to many adaptive finite element algorithms, but where the SOLVE step involves only a full solve on the coarsest level, and the remaining levels involve only single Newton updates to the previous approximate solution. We summarize a recently developed AFEM convergence theory for inexact solvers appearing in [3], and present a sequence of numerical experiments that give evidence that the theory does in fact predict the contraction properties of AFEM with inexact solvers. The various routines used are all designed to maintain a linear-time computational complexity.

An outline of the paper is as follows. In Sect. 2, we give a brief overview of the Poisson-Boltzmann equation. In Sect. 3, we describe AFEM algorithms, and introduce a variation involving inexact solvers. In Sect. 4, we give a sequence of numerical experiments that support the theoretical statements on convergence and optimality. Finally, in Sect. 5 we make some final observations.

2 Regularized Poisson-Boltzmann Equation

We use standard notation for Sobolev spaces. In particular, we denote $\|\cdot\|_{0,G}$ the L^2 norm on any subset $G \subset \mathbb{R}^3$, and denote $\|\cdot\|_{1,2,G}$ the H^1 norm on G.

R. Bank et al. (eds.), *Domain Decomposition Methods in Science and Engineering XX*, Lecture Notes in Computational Science and Engineering 91,
DOI 10.1007/978-3-642-35275-1_18,

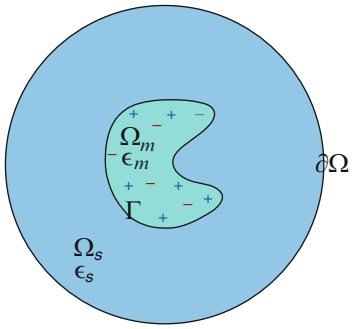


Fig. 1. Schematic of a molecular domain

Let $\Omega := \Omega_m \cup \Gamma \cup \Omega_s$ be a bounded Lipschitz domain in $\mathbb{R}^3$, which consists of the molecular region Ω_m, the solvent region Ω_s and their interface $\Gamma := \overline{\Omega}_m \cap \overline{\Omega}_s$ (see Fig. 1). Our interest in this paper is to solve the following regularized Poisson-Boltzmann equation in the weak form: find $u \in H_g^1(\Omega) := \{u \in H^1(\Omega) : u|_{\partial\Omega} = g\}$ such that

$$a(u,v) + (b(u),v) = (f,v) \quad \forall v \in H_0^1(\Omega), \tag{1}$$

where $a(u,v) = \int_\Omega \varepsilon \nabla u \cdot \nabla v dx$, $(b(u),v) = \int_\Omega \kappa^2 \sinh(u) v dx$. Here we assume that the diffusion coefficient ε is piecewise positive constant $\varepsilon|_{\Omega_m} = \varepsilon_m$ and $\varepsilon|_{\Omega_s} = \varepsilon_s$. The modified Debye-Hückel parameter κ^2 is also piecewise constant with $\kappa^2(x)|_{\Omega_m} = 0$ and $\kappa^2(x)|_{\Omega_s} > 0$. The equation (1) arises from several regularization schemes (cf. [5, 6]) of the nonlinear Poisson-Boltzmann equation:

$$-\nabla \cdot (\varepsilon \nabla u) + \kappa^2 \sinh u = \sum_{i=1}^{N} z_i \delta(x_i),$$

where the right hand side represents N fixed points with charges z_i at positions x_i, and δ is the Dirac delta distribution.

It is easy to verify that the bilinear form in (1) satisfies:

$$c_0 \|u\|_{1,2}^2 \le a(u,u), \qquad a(u,v) \le c_1 \|u\|_{1,2} \|v\|_{1,2}, \qquad \forall u,v \in H_0^1(\Omega),$$

where $0 < c_0 \le c_1 < \infty$ are constants depending only on ε. These properties imply the norm on $H_0^1(\Omega)$ is equivalent to the energy norm $|||\cdot||| : H_0^1(\Omega) \to \mathbb{R}$,

$$|||u|||^2 = a(u,u), \qquad c_0 \|u\|_{1,2}^2 \le |||u|||^2 \le c_1 \|u\|_{1,2}^2.$$

Let $\mathcal{T}_h$ be a shape-regular conforming triangulation of Ω, and let $V_g(\mathcal{T}_h) := \{v \in H_g^1(\Omega) : v|_\tau \in \mathbb{P}_1(\tau) \ \forall \tau \in \mathcal{T}_h\}$ be the standard piecewise linear finite element space defined on $\mathcal{T}_h$. For simplicity, we assume that the interface Γ is resolved by $\mathcal{T}_h$. Then the finite element approximation of (1) reads: find $u_h \in V_g(\mathcal{T}_h)$ such that

$$a(u_h,v) + (b(u_h),v) = (f,v), \quad \forall v \in V_0(\mathcal{T}_h). \tag{2}$$

We close this section with a summary of a priori L^∞ bounds for the solution u to (1) and the discrete solution u_h to (2), which play a key role in the finite element error analysis of (2) and adaptive algorithms. For interested reader, we refer to [5, 9] for details.

Theorem 1. *There exist $u_+, u_- \in L^\infty(\Omega)$ such that the solution u of (1) satisfies the following a priori L^∞ bounds:*

$$u_- \le u \le u_+, \qquad a.e.\ in\ \Omega. \tag{3}$$

Moreover, if the triangulation $\mathscr{T}_h$ satisfies that

$$a(\phi_i, \phi_j) \le -\frac{\sigma}{h^2} \sum_{e_{i,j} \subset \tau} |\tau|, \quad for\ some \quad \sigma > 0, \tag{4}$$

for all the adjacent vertices $i \neq j$ with the basis function ϕ_i and ϕ_j, then the discrete solution u_h of (2) also has the a priori L^∞ bound

$$\|u_h\|_{L^\infty(\Omega)} \le C, \tag{5}$$

where C is a constant independent of h.

We note that the mesh condition is generally not needed practically, and in fact can also be avoided in analysis for certain nonlinearites [2].

3 Adaptive FEM with Inexact Solvers

Given a discrete solution $u_h \in V_g(\mathscr{T}_h)$, let us define the residual based error indicator $\eta(u_h, \tau)$:

$$\eta^2(u_h, \tau) = h_\tau^2 \|b(u_h) - f\|_{0,\tau}^2 + \sum_{e \subset \partial\tau} h_e \|[(\varepsilon \nabla u_h) \cdot n_e]\|_{0,e}^2,$$

where $[(\varepsilon \nabla u_h) \cdot n_e]$ denote the jump of the flux across a face e of τ. For any subset $\mathscr{S} \subset \mathscr{T}_h$, we set $\eta^2(u_h, \mathscr{S}) := \sum_{\tau \in \mathscr{S}} \eta^2(u_h, \tau)$. By using the a priori L^∞ bounds Theorem 1, we can show (cf. [9]) that the error indicator satisfies:

$$|||u - u_h|||^2 \le C_1 \eta^2(u_h, \hat{\mathscr{T}}_h); \tag{6}$$

and

$$|\eta(v, \tau) - \eta(w, \tau)| \le C_2 |||v - w|||_{\omega_\tau}, \quad \forall v, w \in V_g(\mathscr{T}_h) \tag{7}$$

where $\omega_\tau = \cup_{\tau' \in \mathscr{T}_h, \bar{\tau}' \cap \bar{\tau} \neq \emptyset} \tau'$ and $|||v|||_{\omega_\tau}^2 = \int_{\omega_\tau} \varepsilon |\nabla v|^2 dx$.

Given an initial triangulation $\mathscr{T}_0$, the standard adaptive finite element method (AFEM) generates a sequence $\left[u_k, \mathscr{T}_k, \{\eta(u_k, \tau)\}_{\tau \in \mathscr{T}_k}\right]$ based on the iteration of the form:

$$\mathsf{SOLVE} \rightarrow \mathsf{ESTIMATE} \rightarrow \mathsf{MARK} \rightarrow \mathsf{REFINE}.$$

Here the SOLVE subroutine is usually assumed to be exact, namely u_k is the exact solution to the nonlinear equation (2); the ESTIMATE routine computes the element-wise residual indicator $\eta(u_k, \tau)$; the MARK routine uses standard Dörfler marking (cf. [7]) where $\mathscr{M}_k \subset \mathscr{T}_k$ is chosen so that

$$\eta(u_k, \mathscr{M}_k) \geq \theta \eta(u_k, \mathscr{T}_k)$$

for some parameter $\theta \in (0,1]$; finally, the routine REFINE subdivide the marked elements and possibly some neighboring elements in certain way such that the new triangulation preserves shape-regularity and conformity.

During last decade, a lot of theoretical work has been done to show the convergence of the AFEM with exact solver (see [11] and the references cited therein for linear PDE case, and [10] for nonlinear PDE case). To the best of the authors knowledge, there are only a couple of convergence results of AFEM for symmetric linear elliptic equations (cf. [1, 12]) which take the numerical error into account. To distinct with the exact solver case, we use $\hat{u}_k$ and $\hat{\mathscr{T}}_k$ to denote the numerical approximation to (2) and the triangulation obtained from the adaptive refinement using the inexact solutions.

Due to the page limitation, we only state the main convergence result of the AFEM with inexact solver for solving (1) below. More detailed analysis and extension are reported in [3].

Theorem 2. *Let $\{\hat{\mathscr{T}}_k, \hat{u}_k\}_{k\geq 0}$ be the sequence of meshes and approximate solutions computed by the AFEM algorithm. Let u denote the exact solution and u_k denote the exact discrete solutions on the meshes $\hat{\mathscr{T}}_k$. Then, there exist constants $\mu > 0$, $\nu \in (0,1)$, $\gamma > 0$, and $\alpha \in (0,1)$ such that if the inexact solutions satisfy*

$$\mu |||u_k - \hat{u}_k|||^2 + |||u_{k+1} - \hat{u}_{k+1}|||^2 \leq \nu \eta^2(\hat{u}_k, \hat{\mathscr{T}}_k) \tag{8}$$

then

$$|||u - u_{k+1}|||^2 + \gamma \eta^2(\hat{u}_{k+1}, \hat{\mathscr{T}}_{k+1}) \leq \alpha^2(|||u - u_k|||^2 + \gamma \eta^2(\hat{u}_k, \hat{\mathscr{T}}_k)). \tag{9}$$

Consequently, $\lim_{k\to\infty} u_k = \lim_{k\to\infty} \hat{u}_k = u$.

The proof of this theorem is based on the upper bound (6) of the exact solution, the Lipschitz property (7) of the error indicator, Dörfler marking, and the following quasi-orthogonality between the exact solutions:

$$|||u - u_{k+1}|||^2 \leq \Lambda |||u - u_k|||^2 - |||u_{k+1} - u_k|||^2 \tag{10}$$

where Λ can be made close to 1 by refinement. For a proof of the inequality (10), see for example [9].

To achieve the optimal computational complexity, we should avoid solving the nonlinear system (2) as much as we could. The two-grid algorithm [13] shows that a nonlinear solver on a coarse grid combined with a Newton update on the fine grid still yield quasi-optimal approximation. Motivated by this idea, we propose the following AFEM algorithm with inexact solver, which contains only one nonlinear solver on the coarsest grid, and Newton updates on each follow-up steps: In Algorithm 1,

Algorithm 1 : $\left[\hat{u}_k, \hat{\mathscr{T}}_k, \{\eta(\hat{u}_k,\tau)\}_{\tau\in\hat{\mathscr{T}}_k}\right] := \mathsf{Inexact\_AFEM}(\mathscr{T}_0,\theta)$

$\hat{u}_0 = u_0 := \mathsf{NSOLVE}(\mathscr{T}_0)$ *%Nonlinear solver on initial triangulation*
for $k := 0,1,\cdots$ **do**
  $\{\eta(\hat{u}_k,\tau)\}_{\tau\in\hat{\mathscr{T}}_k} := \mathsf{ESTIMATE}(\hat{u}_k,\hat{\mathscr{T}}_k)$
  $\mathscr{M}_k := \mathsf{MARK}(\{\eta(\hat{u}_k,\tau)\}_{\tau\in\hat{\mathscr{T}}_k},\hat{\mathscr{T}}_k,\theta)$
  $\hat{\mathscr{T}}_{k+1} := \mathsf{REFINE}(\hat{\mathscr{T}}_k,\mathscr{M}_k)$
  $\hat{u}_{k+1} := \mathsf{UPDATE}(\hat{u}_k,\hat{\mathscr{T}}_{k+1})$ *%One-step Newton update*
end

the NSOLVE routine is used only on the coarsest mesh and is implemented using Newton's method run to certain convergence tolerance. For the rest of the solutions, a single step of Newton's method is used to update the previous approximation. That is, UPDATE computes $\hat{u}_{k+1}$ such that

$$a(\hat{u}_{k+1}-\hat{u}_k,\phi)+(b'(\hat{u}_k)(\hat{u}_{k+1}-\hat{u}_k),\phi)=0 \tag{11}$$

for every $\phi\in V(\hat{\mathscr{T}}_{k+1})$. We remark that since (11) is only a linear problem, we could use the local multilevel method to solve it in (near) optimal complexity (cf. [4]). Therefore, the overall computational complexity of the Algorithm 1 is nearly optimal.

We should point out that it is not obvious how to enforce the required approximation property (8) that $\hat{u}_k$ must satisfy for the theorem. This is examined in more detail in [3]. However, numerical evidence in the following section shows Algorithm 1 is an efficient algorithm, and the results matches the ones from AFEM with exact solver.

4 Numerical Experiments

In this section we present some numerical experiments to illustrate the result in Theorem 2, implemented with FETK [8]. The software utilizes the standard piecewise-linear finite element space for discretizing (1). Algorithm 1 is implemented with care taken to guarantee that each of the steps runs in linear time relative to the number of vertices in the mesh. The linear solver used is Multigrid preconditioned Conjugate Gradients. The estimator is computed using a high-order quadrature rule, and, as mentioned above, the marking strategy is Dörfler marking where the estimated errors have been binned to maintain linear complexity while still marking the elements with the largest error. Finally, the refinement is longest edge bisection, with refinement outside of the marked set to maintain conformity of the mesh.

We present two sets of results in order to explore the effects of the inexact solver in multiple contexts. For each problem, we present a convergence plot using both inexact and exact solvers (including a reference line of order $N^{-\frac{1}{3}}$) as well as a representative cut-away of a mesh with around 30,000 vertices. The exact discrete solution is computed using the standard AFEM algorithm where the solution on each

mesh is computed by allowing Newton's method to continue running to convergence with the tolerance 10^{-7}. For the exact solution, one could choose to start with an arbitrary initial guess, such as the zero solution, or, as we've chosen, use the solution computed on the previous mesh. Making this choice can drastically decrease the number of Newton steps needed to achieve convergence. For each problem below, we discuss the amount of time/computation saved using the inexact solver over this exact solver.

Note that using the inexact solver modifies not only the solution on a given mesh, but also the sequence of meshes generated, since the algorithm may mark different simplices. However, as shown in the examples below, the inexact solutions still maintain optimal convergence rates.

The first result uses constant coefficients across the entire domain $\Omega = [0,1]^3$, an exponential nonlinearity, and a right hand side chosen so that the derivative of the exact solution is large near the origin. The boundary conditions chosen for this problem are homogeneous Dirichlet boundary conditions. Specifically, the exact solution is given by $u = u_1 u_2$ where

$$u_1 = \sin(\pi x)\sin(\pi y)\sin(\pi z)$$

is chosen to satisfy the boundary condition and

$$u_2 = 3(x^2 + y^2 + x^2 + 10^{-4})^{-1.5}.$$

The results can be seen in Fig. 2.

For this problem, the number of iterations in Newton's method by the exact solver varied between 3 and 7, depending on the refinement level. Because all steps of the algorithm are designed to be linear, this suggests that the inexact solver runs at least three times faster for this problem, while still maintaining optimal order of convergence.

In order to test the robustness to the addition of jump coefficients, the second result uses the domain $\Omega = [-1,1]^3$ and $\Omega_m = \left[-\frac{1}{4}, \frac{1}{4}\right]$ with constants $\varepsilon_s = 80, \varepsilon_m = 2, \kappa_s = 1$, and $\kappa_m = 0$. Homogeneous Neumann conditions are chosen for the boundary and the right hand side is simplified to a constant. Because an exact solution is unavailable for this (and the following) problem, the error is computed by comparing to a discrete solution on a mesh with around ten times the number of vertices as the finest mesh used in the adaptive algorithm. Figure 3 shows the results for this problem. As can be seen the refinement favors the interface and the inexact and exact solvers perform as expected.

Once again, for this problem, the exact solver required between 3 and 9 iterations of Newton's method to reach convergence, depending on the refinement level. Since the run time is linear is the number of iterations, this result gives a speedup of at least three times using the inexact solver, without causing a loss in convergence rate.

5 Conclusion

In this article we have studied AFEM with inexact solvers for a class of semilinear elliptic interface problems with discontinuous diffusion coefficients. The algorithm

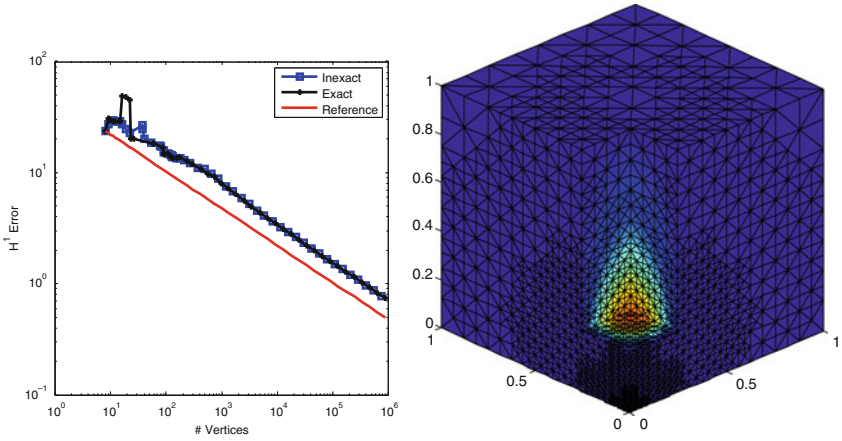


Fig. 2. Convergence plot and mesh cut-away for the corner singularity problem

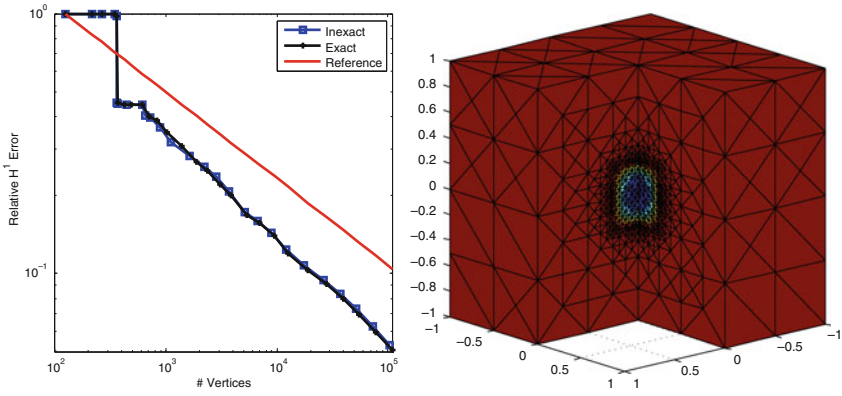


Fig. 3. Convergence plot and mesh cut-away for the Poisson-Boltzmann problem

we studied consisted of the standard SOLVE-ESTIMATE-MARK-REFINE procedure common to many adaptive finite element algorithms, but where the SOLVE step involves only a full solve on the coarsest level, and the remaining levels involve only single Newton updates to the previous approximate solution. Our numerical results indicate that the recently developed AFEM convergence theory for inexact solvers in [3] does predict the actual behavior of the methods and can allow for significant speedup in the approximation of solutions.

Bibliography

[1] M. Arioli, E.H. Georgoulis, and D. Loghin. Convergence of inexact adaptive finite element solvers for elliptic problems. Technical Report RAL-TR-2009-021, Science and Technology Facilities Council, October 2009.

[2] R. Bank, M. Holst, R. Szypowski, and Y. Zhu. Finite element error estimates for critical exponent semilinear problems without mesh conditions. Preprint, 2011.

[3] R. Bank, M. Holst, R. Szypowski, and Y. Zhu. Convergence of AFEM for semilinear problems with inexact solvers. Preprint, 2011.

[4] L. Chen, M. Holst, J. Xu, and Y. Zhu. Local Multilevel Preconditioners for Elliptic Equations with Jump Coefficients on Bisection Grids. *Arxiv preprint arXiv:1006.3277*, 2010.

[5] L. Chen, M. Holst, and J. Xu. The finite element approximation of the nonlinear Poisson-Boltzmann equation. *SIAM Journal on Numerical Analysis*, 45(6): 2298–2320, 2007.

[6] I-L. Chern, J.-G. Liu, and W.-C. Wan. Accurate evaluation of electrostatics for macromolecules in solution. *Methods and Applications of Analysis*, 10: 309–328, 2003.

[7] W. Dörfler. A convergent adaptive algorithm for Poisson's equation. *SIAM Journal on Numerical Analysis*, 33:1106–1124, 1996.

[8] FETK. The Finite Element ToolKit. http://www.FETK.org.

[9] M. Holst, J.A. McCammon, Z. Yu, Y.C. Zhou, and Y. Zhu. Adaptive Finite Element Modeling Techniques for the Poisson-Boltzmann Equation. *Accepted for publication in Communications in Computational Physics*, 2009.

[10] M. Holst, G. Tsogtgerel, and Y. Zhu. Local Convergence of Adaptive Methods for Nonlinear Partial Differential Equations. *arXiv*, (1001.1382v1), 2010.

[11] R.H. Nochetto, K.G. Siebert, and A. Veeser. Theory of adaptive finite element methods: An introduction. In R.A. DeVore and A. Kunoth, editors, *Multiscale, Nonlinear and Adaptive Approximation*, pages 409–542. Springer, 2009.

[12] R. Stevenson. Optimality of a standard adaptive finite element method. *Found. Comput. Math.*, 7(2):245–269, 2007.

[13] J. Xu. Two-grid discretization techniques for linear and nonlinear PDEs. *SIAM Journal on Numerical Analysis*, 33(5):1759–1777, 1996.

Preconditioning for Mixed Finite Element Formulations of Elliptic Problems

Tim Wildey[1] and Guangri Xue[2]

[1] Sandia National Labs, Albuquerque, NM 87185 (tmwilde@sandia.gov). Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

[2] The University of Texas at Austin, Institute for Computational Engineering and Sciences, 201 E 24th Street, Austin TX 78712, USA, gxue@ices.utexas.edu (current address: Shell International Exploration and Production Inc., 3737 Bellaire Blvd, Rm 2009, Houston, TX 77025, USA, Guangri.Xue@shell.com).

Summary. In this paper, we discuss a preconditioning technique for mixed finite element discretizations of elliptic equations. The technique is based on a block-diagonal approximation of the mass matrix which maintains the sparsity and positive definiteness of the corresponding Schur complement. This preconditioner arises from the multipoint flux mixed finite element method and is robust with respect to mesh size and is better conditioned for full permeability tensors than a preconditioner based on a diagonal approximation of the mass matrix.

1 Introduction

Consider the mixed formulation of a second order linear elliptic equation. Introducing a flux variable, we solve for a scalar potential p and a vector function $\mathbf{u}$ that satisfy

$$\mathbf{u} = -\mathbb{K}\nabla p \quad \text{in } \Omega, \tag{1}$$

$$\nabla \cdot \mathbf{u} = f \quad \text{in } \Omega, \tag{2}$$

$$p = 0 \quad \text{on } \partial\Omega, \tag{3}$$

where Ω is a polygonal domain with Lipschitz continuous boundary and $\mathbb{K}$ is a symmetric and uniformly positive definite tensor with $L^\infty(\Omega)$ components. Homogeneous Dirichlet boundary conditions are considered for the simplicity of the presentation.

Mixed finite element methods lead to the non-singular indefinite system:

$$\mathbb{M}\begin{pmatrix} U \\ P \end{pmatrix} := \begin{pmatrix} \mathbb{A} & \mathbb{B}^T \\ \mathbb{B} & 0 \end{pmatrix}\begin{pmatrix} U \\ P \end{pmatrix} = \begin{pmatrix} 0 \\ F \end{pmatrix}, \tag{4}$$

where the matrix $\mathbb{A}$ is a symmetric and positive definite.

R. Bank et al. (eds.), *Domain Decomposition Methods in Science and Engineering XX*, Lecture Notes in Computational Science and Engineering 91,
DOI 10.1007/978-3-642-35275-1_19,

In this paper, we consider preconditioners of the form:

$$\widetilde{\mathbb{M}} := \begin{pmatrix} \widetilde{\mathbb{A}} & \mathbb{B}^T \\ \mathbb{B} & 0 \end{pmatrix}. \tag{5}$$

The applicability of this type preconditioner is due to the fact that

- $\widetilde{\mathbb{A}}$ is easily invertible.
- The Schur complement of the preconditioner $\widetilde{M}$ is sparse and positive definite, and can be solved easily.

One way is choosing $\widetilde{\mathbb{A}}$ as a diagonal matrix. In [1], $\widetilde{\mathbb{A}}$ is given as $\omega\mathbb{I}$. The global parameter ω is chosen to minimize the spectral radius of $\mathbb{I} - \widetilde{\mathbb{M}}^{-1}\mathbb{M}$. In [5], the diagonal matrix is optimally scaled at element level and a precise upper bound of the spectral radius has been shown: $\rho(\mathbb{I} - \widetilde{\mathbb{M}}^{-1}\mathbb{M}) \leq 1/2$. In other words, the preconditioner is independent of both the mesh size and the tensor $\mathbb{K}$. This uniformity is derived when the problem has a diagonal $\mathbb{K}$ and is discretized by the lowest order Raviart-Thomas [8] mixed finite element on rectangular grids. For other mixed finite element spaces or full tensor $\mathbb{K}$, the uniformity result is not clearly understood. Alternatively, a simple parameter-free choice for $\widetilde{\mathbb{A}}$, $\widetilde{\mathbb{A}} = \mathrm{Diag}(\mathbb{A})$, can be used.

Another approach is to take $\widetilde{\mathbb{A}}$ as a block-diagonal matrix which guarantees that the corresponding Schur complement matrix is sparse and positive definite. Multipoint flux mixed finite element (MFMFE) methods [6, 9–12] give matrices of the form (5), where the flux variable can be locally eliminated due to the block-diagonal structure of $\widetilde{\mathbb{A}}$. The corresponding Schur complement gives a cell-centered stencil for the scalar variable. In this paper, we study the performance of this MFMFE operator as a preconditioner. The Schur complement of MFMFE has a 9-point stencil on logically rectangular grids and with full tensor $\mathbb{K}$ in contrast to 5-point stencil which arises if $\widetilde{\mathbb{A}}$ is a diagonal matrix. Our numerical result indicates that the MFMFE method gives a better preconditioner than the diagonal preconditioner ($\widetilde{\mathbb{A}} = \mathrm{Diag}(\mathbb{A})$). A natural extension of this work is the use of approximate preconditioners based on algebraic multigrid for MFMFE as described in [2, 7] and will be the subject of future work.

The rest of the paper is organized as follows. Mixed finite element formulation is described in Sect. 2. A block type preconditioner is discussed in Sect. 3. Finally in Sect. 4, numerical experiments are given.

2 Mixed Finite Element Formulation

Define $H(\mathrm{div};\Omega) := \left\{\mathbf{v} \in (L^2(\Omega))^d : \nabla\cdot\mathbf{v} \in L^2(\Omega)\right\}$ and let $(\cdot,\cdot)$ denote the inner product in $L^2(\Omega)$. Let $X \lesssim (\gtrsim)\, Y$ denote that there exists a constant C, independent of the mesh size h, such that $X \leq (\geq)\, CY$. The notation $X \eqsim Y$ means that both $X \lesssim Y$ and $X \gtrsim Y$ hold.

Let $\mathscr{T}_h$ be a finite element partition of the domain Ω consisting of either triangles or quadrilaterals. We assume that $\mathscr{T}_h$ is shape-regular in the sense of Ciarlet [4].

The finite element spaces on any physical element $E \in \mathscr{T}_h$ are defined via the Piola transformation

$$\mathbf{v} \leftrightarrow \hat{\mathbf{v}} : \hat{\mathbf{v}} = \frac{1}{J_E} \mathbb{DF}_E \hat{\mathbf{v}} \circ F_E^{-1},$$

and the scalar transformation

$$w \leftrightarrow \hat{w} : w = \hat{w} \circ F_E^{-1},$$

where F_E denotes a mapping from the reference element $\hat{E}$ to the physical element E, $\mathbb{DF}_E$ is the Jacobian of F_E, and J_E is its determinant. The finite element spaces V_h and W_h on $\mathscr{T}_h$ are given by

$$V_h = \left\{ \mathbf{v} \in H(\mathrm{div}; \Omega) : \quad \mathbf{v}|_E \leftrightarrow \hat{\mathbf{v}},\ \hat{\mathbf{v}} \in \hat{V}(\hat{E}), \quad \forall E \in \mathscr{T}_h \right\},$$
$$W_h = \left\{ w \in L^2(\Omega) : \quad w|_E \leftrightarrow \hat{w},\ \hat{w} \in \hat{W}(\hat{E}), \quad \forall E \in \mathscr{T}_h \right\},$$

where $V(\hat{E})$ and $\hat{W}(\hat{E})$ are the lowest order Brezzi-Douglas-Marini (BDM_1) spaces on the reference element $\hat{E}$. Definitions of Piola transformation and BDM_1 spaces yield $V_h \subset H(\mathrm{div}; \Omega)$ and $W_h \subset L^2(\Omega)$.

The finite element method reads: find $\mathbf{u}_h \in V_h$ and $p_h \in W_h$, such that

$$(\mathbb{K}^{-1}\mathbf{u}_h, \mathbf{v}) - (p_h, \nabla \cdot \mathbf{v}) = 0, \qquad \forall \mathbf{v} \in V_h, \tag{6}$$
$$-(\nabla \cdot \mathbf{u}_h, w) = -(f, w) \quad \forall w \in W_h. \tag{7}$$

The method (6) and (7) can have a second order convergence for the flux and first order convergence for the scalar potential [3] if $\mathbf{u}$ and p are sufficiently regular.

3 Preconditioning the Mixed Finite Element System

3.1 Multipoint Flux Mixed Finite Element

A family of multipoint flux mixed finite element (MFMFE) methods on various grids has been developed and analyzed [6, 9–12]. The method is defined as: find $\mathbf{u}_h \in V_h$ and $p_h \in W_h$, such that

$$(\mathbb{K}^{-1}\mathbf{u}_h, \mathbf{v})_Q - (p_h, \nabla \cdot \mathbf{v}) = 0, \qquad \forall \mathbf{v} \in V_h, \tag{8}$$
$$-(\nabla \cdot \mathbf{u}_h, w) = -(f, w) \quad \forall w \in W_h, \tag{9}$$

where the finite element spaces are BDM_1 on triangular and rectangular meshes. Compared to the BDM_1 finite element method, a specific numerical quadrature rule is employed. It is defined as:

$$(\mathbb{K}^{-1}\mathbf{q}, \mathbf{v})_Q = \sum_{E \in \mathscr{T}_h} (\mathbb{K}^{-1}\mathbf{q}, \mathbf{v})_{Q,E} \equiv \sum_{E \in \mathscr{T}_h} \mathrm{Trap}\left(\mathscr{K}\hat{\mathbf{q}}, \hat{\mathbf{v}}\right)_{\hat{E}}, \tag{10}$$

where $\mathscr{K}$ on each $\hat{E}$ is defined as

$$\mathscr{K} = \frac{1}{J_E}\mathbb{D}\mathbb{F}_E^T\mathbb{K}^{-1}(F_E(\hat{x}))\mathbb{D}\mathbb{F}_E, \tag{11}$$

and the trapezoidal rule on $\hat{E}$ is denoted as

$$\text{Trap}(\hat{\mathbf{q}},\hat{\mathbf{v}})_{\hat{E}} \equiv \frac{|\hat{E}|}{m}\sum_{i=1}^{m}\hat{\mathbf{q}}(\hat{\mathbf{r}}_i)\cdot\hat{\mathbf{v}}(\hat{\mathbf{r}}_i), \tag{12}$$

with $\{\hat{\mathbf{r}}_i\}_{i=1}^m$ being vertices of $\hat{E}$ and m being the number of vertices of $\hat{E}$.

The degrees of freedom for the flux variable are chosen as the normal components at two vertices on each edge. More specifically, denote the basis functions associated with $\hat{\mathbf{r}}_i$ by $\hat{\mathbf{v}}_{ij}, j = 1,2$: $(\hat{\mathbf{v}}_{ij}\cdot\hat{\mathbf{n}}_{ij})(\hat{\mathbf{r}}_i) = 1$, $(\hat{\mathbf{v}}_{ij}\cdot\hat{\mathbf{n}}_{ik})(\hat{\mathbf{r}}_i) = 0, k \neq j$, and $(\hat{\mathbf{v}}_{ij}\cdot\hat{\mathbf{n}}_{lk})(\hat{\mathbf{r}}_l) = 0, l \neq i$, k = 1, 2. As a consequence, the quadrature rule (10) couples only the two basis functions associated with a vertex. For example, on the unit square

$$\begin{aligned}
&(\mathscr{K}\hat{\mathbf{v}}_{11},\hat{\mathbf{v}}_{11})_{\hat{Q},\hat{E}} = \frac{\mathscr{K}_{11}(\hat{\mathbf{r}}_1)}{4}, \quad (\mathscr{K}\hat{\mathbf{v}}_{11},\hat{\mathbf{v}}_{12})_{\hat{Q},\hat{E}} = \frac{\mathscr{K}_{21}(\hat{\mathbf{r}}_1)}{4},\\
&(\mathscr{K}\hat{\mathbf{v}}_{11},\hat{\mathbf{v}}_{ij})_{\hat{Q},\hat{E}} = 0, \;\; i \neq 1, j = 1,2.
\end{aligned} \tag{13}$$

where $\mathscr{K}_{ij}$ denotes i-th row and j-th column of the matrix function $\mathscr{K}$. This localization property on interactions between the flux basis functions gives the assembled mass matrix in (8) has a block diagonal structure with one block per grid vertex.

We denote the algebraic system arising from (8) and (9) as

$$\begin{pmatrix} \mathbb{A}_Q & \mathbb{B}^T \\ \mathbb{B} & 0 \end{pmatrix}\begin{pmatrix} U \\ P \end{pmatrix} = \begin{pmatrix} 0 \\ F \end{pmatrix}, \tag{14}$$

where $\mathbb{A}_Q$ is block diagonal. The approximate flux, U, can be easily eliminated via

$$U = -\mathbb{A}_Q^{-1}\mathbb{B}^T P. \tag{15}$$

The resulting Schur complement system

$$\mathbb{B}\mathbb{A}_Q^{-1}\mathbb{B}^T P = -F, \tag{16}$$

is symmetric positive definite and sparse. On rectangular grids, Eq. (16) has a 5-point stencil for a diagonal tensor $\mathbb{K}$ and 9-point stencil for the full tensor. The Schur complement system can be solved using classical algebraic multigrid methods. The flux variable is then obtained easily by (15) due to the block diagonal structure of $\mathbb{A}_Q$.

The following result concerns the convergence of the MFMFE methods. Let $W^{k,\infty}_{\mathscr{T}_h}$ consist of functions ϕ such that $\phi|_E \in W^{k,\infty}(E)$ for all $E \in \mathscr{T}_h$.

Theorem 1 ([6, 10–12]). *Let $\mathscr{T}_h$ consist of simplices, h^2-parallelograms, h^2-parallelepipeds or triangular prisms. If $\mathbb{K}^{-1} \in W^{1,\infty}_{\mathscr{T}_h}$, then, the flux $\mathbf{u}_h$ and scalar p_h of the MFMFE method (8)–(9) satisfies*

$$\|\mathbf{u}-\mathbf{u}_h\| \lesssim h\|\mathbf{u}\|_1, \;\; \|\nabla\cdot(\mathbf{u}-\mathbf{u}_h)\| \lesssim h\|\nabla\cdot\mathbf{u}\|_1, \;\; \|p-p_h\| \lesssim h(\|\mathbf{u}\|_1 + \|p\|_1).$$

Compared to the second order L^2 convergence of the flux variable in the BDM_1 mixed method, the MFMFE has a first order convergence for the flux variable due to the numerical quadrature. However the MFMFE method is a solver friendly scheme since the MFMFE method can be reduced to a cell-centered stencil in terms of the scalar variable without solving a saddle-point problem.

3.2 Multipoint Flux Mixed Finite Element as a Preconditioner

The MFMFE method may be used as a preconditioner to the BDM_1 mixed finite element method by choosing $\widetilde{\mathbb{A}} = \mathbb{A}_Q$.

Lemma 1. *The condition number of $\widetilde{\mathbb{A}}^{-1}\mathbb{A}$ is independent of the mesh size.*

Proof. It has been shown [6, 11, 12] that the bilinear form $(\mathbb{K}^{-1}\cdot,\cdot)_Q$ is an inner product in $\mathbf{V}_h$ and $(\mathbb{K}^{-1}\mathbf{q},\mathbf{q})_Q^{1/2}$ is a norm equivalent to the L^2 norm. Thus

$$(\mathbb{K}^{-1}\mathbf{q},\mathbf{q})_Q \eqsim \|\mathbf{q}\|^2 \eqsim (\mathbb{K}^{-1}\mathbf{q},\mathbf{q}), \quad \forall \mathbf{q} \in \mathbf{V}_h. \qquad \square \tag{17}$$

The preconditioner of the form (5) has been analyzed by Ewing, Lazarov, Lu and Vassilevski.

Theorem 2 ([5]). *The eigenvalues of $\widetilde{\mathbb{M}}^{-1}\mathbb{M}$ are real and positive and lie in the interval $[\lambda_{min}, \lambda_{max}]$, where λ_{min} and λ_{max} are the extreme eigenvalues of $\widetilde{\mathbb{A}}^{-1}\mathbb{A}$.*

By Lemma 1 and Theorem 2, we have the following corollary.

Corollary 1. *The preconditioned system of BDM_1 mixed finite element method with MFMFE as a preconditioner is positive definite. The condition number is independent of the mesh size.*

4 Numerical Results

4.1 Example 1

In this example, we consider (1)–(3) on the computational domain shown in Fig. 1 (left) with $p = 0$ on $\partial\Omega$ and $f = 1$.

First, we use the MFMFE method as a preconditioner for the BDM_1 mixed finite element method with $\mathbb{K} = \mathbb{I}$. The result is presented in Table 1 where we can clearly see that the preconditioner is robust with respect to the mesh size h. Next, we consider the heterogeneous permeability field shown in Fig. 1 (right) which is generated using geostatistical techniques (kriging) with a longer correlation length in the horizontal direction. In Table 2 we see that the preconditioner is not only robust with respect to mesh size, but also with respect to the heterogeneities in the permeability.

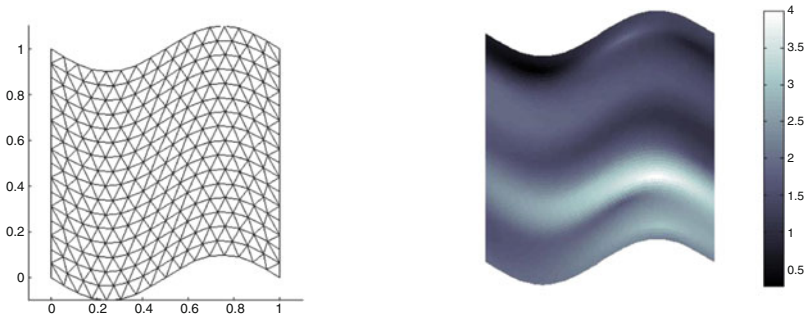


Fig. 1. The triangular mesh used in Example 1 with $h \approx 1/16$ (*left*) and the log of the heterogeneous permeability field (*right*)

h	Degrees of Freedom	$\text{cond}(\widetilde{\mathbb{M}}^{-1}\mathbb{M})$
$1/8$	512	13.43
$1/16$	2048	15.84
$1/32$	8192	15.61
$1/64$	32768	15.63

Table 1. Performance of the MFMFE preconditioner with a homogeneous permeability field.

h	Degrees of Freedom	$\text{cond}(\widetilde{\mathbb{M}}^{-1}\mathbb{M})$
$1/8$	512	20.07
$1/16$	2048	21.61
$1/32$	8192	16.61
$1/64$	32768	14.27

Table 2. Performance of the MFMFE preconditioner with a heterogeneous permeability field.

4.2 Example 2

In this example, we consider (1)–(3) with $\Omega = [0,1] \times [0,1]$ and

$$\mathbb{K} = \begin{pmatrix} 1+\alpha & 1-\alpha \\ 1-\alpha & 1+\alpha \end{pmatrix},$$

with $0 < \alpha \leq 1$. We use uniform rectangular meshes and our objective is to demonstrate that the MFMFE preconditioner is more robust as $\alpha \to 0$. In Tables 3 and 4 we present the results using the diagonal preconditioner ($\widetilde{\mathbb{A}} = \text{Diag}(\mathbb{A})$) and the MFMFE preconditioner respectively. We see that both preconditioners are robust with respect to h, but degrade as $\alpha \to 0$, but the MFMFE preconditioner degrades at a much slower rate.

α	$h = 1/4$	$h = 1/8$	$h = 1/16$	$h = 1/32$
1	22.43	22.32	22.32	22.32
1E-1	1.06E2	9.95E2	1.06E2	1.06E2
1E-2	7.00E2	6.97E2	6.97E2	6.97E2
1E-3	9.51E3	9.41E3	9.75E3	8.42E3

Table 3. Performance of a diagonal preconditioner with respect to h and α.

α	$h = 1/4$	$h = 1/8$	$h = 1/16$	$h = 1/32$
1	22.42	22.32	22.32	22.32
1E-1	32.07	32.09	32.26	32.09
1E-2	51.01	50.06	50.39	50.39
1E-3	5.20E2	6.96E2	8.10E2	8.21E2

Table 4. Performance of the MFMFE preconditioner with respect to h and α.

5 Conclusions

The purpose of this paper is to investigate the performance of the multipoint flux mixed finite element as a preconditioner for the saddle-point system for the full BDM_1 mixed finite element approximation. Numerical results indicate that the MFMFE preconditioner is robust with respect to the mesh size and performs better than the preconditioner based on the diagonal mass matrix.

Acknowledgments Guangri Xue is supported by Award No. KUS-F1-032-04, made by King Abdullah University of Science and Technology (KAUST).

Bibliography

[1] M. B. Allen, R. E. Ewing, and P. Lu. Well-conditioned iterative schemes for mixed finite-element models of porous-media flows. *SIAM J. Sci. Stat. Comput.*, 13:794–814, 1992.

[2] M. Benzi, G. H. Golub, and J. Liesen. Numerical solution of saddle point problems. *Acta Numerica*, 14:1–137, 2005.

[3] F. Brezzi, J. Douglas, and L. D. Marini. Two families of mixed finite elements for second order elliptic problems. *Numer. Math.*, 47(2):217–235, 1985.

[4] P. G. Ciarlet. *The Finite Element Method for Elliptic Problems*. Stud. Math. Appl. 4, North-Holland, Amsterdam, 1978; reprinted, SIAM, Philadelphia, 2002.

[5] R. E. Ewing, R. D. Lazarov, P. Lu, and P. Vassilevski. Preconditioning indefinite systems arising from mixed finite element discretization of second-order elliptic problems. *Lecture Notes in Mathematics*, 1457:28–43, 1990.

[6] R. Ingram, M. F. Wheeler, and I. Yotov. A multipoint flux mixed finite element method on hexahedra. *SIAM J. Numer. Anal.*, 48:1281–1312, 2010.

[7] M. F. Murphy, G. H. Golub, and A. J. Wathen. A note on preconditioning for indefinite linear systems. *SIAM J. Sci. Comput.*, 21:1969–1972, 2000.

[8] P. A. Raviart and J. Thomas. A mixed finite element method for 2-nd order elliptic problems. In I. Galligani and E. Magenes, editors, *Mathematical aspects of the Finite Elements Method*, Lectures Notes in Math. 606, pages 292–315. Springer, Berlin, 1977.

[9] M. Wheeler, G. Xue, and I. Yotov. A multipoint flux mixed finite element method on distorted quadrilaterals and hexahedra. *Numer. Math.*, DOI 10.1007/s00211-011-0427-7, 2011.

[10] M. F. Wheeler, G. Xue, and I. Yotov. A family of multipoint flux mixed finite element methods for elliptic problems on general grids. *Procedia Computer Science*, 4:918–927, 2011.

[11] M. F. Wheeler, G. Xue, and I. Yotov. A multipoint flux mixed finite element method on triangular prisms. *Preprint*, 2011.

[12] M. F. Wheeler and I. Yotov. A multipoint flux mixed finite element method. *SIAM. J. Numer. Anal.*, 44(5):2082–2106, 2006.

Multigrid Preconditioner for Nonconforming Discretization of Elliptic Problems with Jump Coefficients

Blanca Ayuso De Dios[1], Michael Holst[2], Yunrong Zhu[3], and Ludmil Zikatanov[4]

[1] Centre de Recerca Matematica (CRM), Barcelona, Spain bayuso@crm.cat
[2] Department of Mathematics, University of California at San Diego, California, USA {mholst, zhu}@math.ucsd.edu
[3] Department of Mathematics, Idaho State University, Pocatello, ID 83209-8085, USA zhuyunr@isu.edu
[4] Department of Mathematics, The Pennsylvania State University, University Park, PA 16802, USA ludmil@psu.edu

Summary. In this paper, we present a multigrid preconditioner for solving the linear system arising from the piecewise linear nonconforming Crouzeix-Raviart discretization of second order elliptic problems with jump coefficients. The preconditioner uses the standard conforming subspaces as coarse spaces. Numerical tests show both robustness with respect to the jump in the coefficient and near-optimality with respect to the number of degrees of freedom.

1 Introduction

The purpose of this paper is to present a multigrid preconditioner for solving the linear system arising from the $\mathbb{P}^1$ nonconforming Crouzeix-Raviart (CR) discretization of second order elliptic problems with jump coefficients. The multigrid preconditioner we consider here uses pointwise relaxation (point Gauss-Seidel/Jacobi iterative methods) as a smoother, followed by a subspace (coarse grid) correction which uses the standard multilevel structure for the nested $\mathbb{P}_1$ conforming finite element spaces. The subspace correction step is motivated by the observation that the standard $\mathbb{P}^1$ conforming space is a subspace of the CR finite element space.

The idea of using conforming subspaces to construct preconditioners for CR discretization has been used in [6, 9, 11] in the context of smooth coefficients. To deal with the jump coefficient problems, multilevel methods using conforming subspaces were proposed and analyzed in [7, 8]. In particular, in [5] it was shown that if the coefficients satisfy the *quasi-monotonicity* condition then the preconditioned systems have condition numbers independent of the coefficients and depending on the mesh size logarithmically. It was also shown in [5] that the same conclusions hold for multilevel preconditioners utilizing a correction in an additional *exotic* coarse space in case of general coefficient distributions with cross points.

R. Bank et al. (eds.), *Domain Decomposition Methods in Science and Engineering XX*, Lecture Notes in Computational Science and Engineering 91,
DOI 10.1007/978-3-642-35275-1_20,

We take a different approach in this paper, and without adding additional coarse space, we show that the multigrid method is a robust preconditioner for PCG algorithm. In particular, we show that the preconditioned system has only a few "bad eigenvalues" (depending on the jumps of the coefficients), and the asymptotic convergence rate of the PCG algorithm is uniform with respect to the coefficient. The analysis follows closely [12] with the help of special technical tools developed in [2]. Due to space limitation we only state the main result (Theorem 1 in Sect. 3), and provide numerical results that support it. Detailed analyses and further discussion of the algorithm are presented in [13]. One of the main benefits of this algorithm is that it is very easy to implement in practice. The procedure is the same as the standard multigrid algorithm on conforming spaces, and the only difference is the prolongation and restriction matrices on the finest level. Since the spaces are nested, the prolongation matrix is simply the matrix representation of the natural inclusion operator from the conforming space to the CR space.

The paper is organized as follows. In Sect. 2, we give basic notation and the finite element discretizations. In Sect. 3, we present the multigrid algorithm and discuss its implementation and convergence. Finally, in Sect. 4 we verify numerically the theoretical results by presenting several numerical tests for two and three dimensional model problems.

2 Preliminaries

Let $\Omega \subset \mathbb{R}^d$ $(d = 2,3)$ be an open polygonal domain. Given $f \in L^2(\Omega)$, we consider the following model problem: Find $u \in H_0^1(\Omega)$ such that

$$a(u,v) := (\kappa \nabla u, \nabla v) = (f,v) \quad \forall v \in H_0^1(\Omega), \tag{1}$$

where the diffusion coefficient $\kappa \in L^\infty(\Omega)$ is assumed to be piecewise constant, namely, $\kappa(x)|_{\Omega_m} = \kappa_m$ is a constant for each (open) polygonal subdomain Ω_m satisfying $\cup_{m=1}^M \overline{\Omega}_m = \overline{\Omega}$ and $\Omega_m \cap \Omega_n = \emptyset$ for $m \neq n$.

We assume that there is an initial (quasi-uniform) triangulation $\mathscr{T}_0$, with mesh size h_0, such that for all $T \in \mathscr{T}_0$ $\kappa_T := \kappa(x)|_T$ is constant. Let $\mathscr{T}_j := \mathscr{T}_{h_j}$ $(j = 1, \cdots, J)$ be a family of uniform refinement of $\mathscr{T}_0$ with mesh size h_j. Without loss of generality, we assume that the mesh size $h_j \simeq 2^{-j} h_0$ $(j = 0, \cdots, J)$ and denote $h = h_J$.

On each level $j = 0, \cdots, J$, we define V_j as the standard $\mathbb{P}^1$ conforming finite element space defined on $\mathscr{T}_j$. Then the standard conforming finite element discretization of (1) reads:

$$\text{Find } u_j \in V_j \text{ such that } a(u_j, v_j) = (f, v_j), \qquad \forall v_j \in V_j. \tag{2}$$

For each $j = 0, \cdots, J$, we define the induced operator for (2) as

$$(A_j v_j, w_j) = a(v_j, w_j), \quad \forall v_j, w_j \in V_j.$$

We denote $\mathscr{E}_h$ the set of all edges (in 2D) or faces (in 3D) of $\mathscr{T}_h$. Let V_h^{CR} be the piecewise linear nonconforming Crouzeix-Raviart finite element space defined by:

$$V_h^{CR} = \left\{ v \in L^2(\Omega) : v_{|_T} \in \mathbb{P}^1(T) \forall T \in \mathscr{T}_h \text{ and } \int_e [\![v]\!]_e ds = 0 \, \forall e \in \mathscr{E}_h \right\},$$

where $\mathbb{P}^1(T)$ denotes the space of linear polynomials on T and $[\![v]\!]_e$ denotes the jump across the edge/face $e \in \mathscr{E}_h$ with $[\![v]\!]_e = v$ when $e \subset \partial\Omega$. In the sequel, let us denote $V_{J+1} := V_h^{CR}$ for simplicity. We remark that all these finite element spaces are nested, that is,

$$V_0 \subset \cdots \subset V_J \subset V_{J+1}.$$

The $\mathbb{P}^1$-nonconforming finite element approximation to (1) reads:

$$\text{Find } u \in V_h^{CR} : a_h(u,w) := \sum_{T \in \mathscr{T}_J} \int_T \kappa_T \nabla u \cdot \nabla w = (f,w), \forall w \in V_h^{CR}. \tag{3}$$

The bilinear form $a_h(\cdot,\cdot)$ induced a natural energy norm: $|v|_{h,\kappa} := \sqrt{a_h(v,v)}$ for any $v \in V_h^{CR}$. In operator form, we are going to solve the linear system

$$Au = f, \tag{4}$$

where A is the operator induced by (3), namely

$$(Av,w) = a_h(v,w), \qquad \forall v,w \in V_h^{CR}.$$

3 A Multigrid Preconditioner

The action of the standard multigrid V-cycle preconditioner $B := B_{J+1} : V_{J+1} \mapsto V_{J+1}$ on a given $g \in V_{J+1}$ is recursively defined by the following algorithm (cf. [3]):

V-cycle

Let $g_{J+1} = g$, and $B_0 = A_0^{-1}$. For $j = 1, \cdots, J+1$, we define recursively $B_j g_j$ for any $g_j \in V_j$ by the following three steps:

1. Pre-smoothing : $w_1 = R_j g_j$;
2. Subspace correction: $w_2 = w_1 + B_{j-1} Q_{j-1}(g_j - A_j w_1)$;
3. Post-smoothing: $B_j g_j := w_2 + R_j^*(g_j - A_j w_2)$.

In this algorithm, R_j corresponds to a Gauss-Seidel or a Jacobi iterative method known as a smoother; and Q_j is the standard L^2 projection on V_j:

$$(Q_j v, w_j) = (v, w_j), \qquad \forall w_j \in V_j, \ \ (j = 0, \cdots, J).$$

The implementation of Algorithm 3 is almost identical to the implementation of the standard multigrid V-cycle (cf. [4]). Between the conforming spaces, we use the standard prolongation and restriction matrices (for conforming finite elements). The

corresponding matrices between V_J and V_{J+1}, are however different. The prolongation matrix on V_J can be viewed as the matrix representation of the natural inclusion $\mathscr{I}_J : V_J \to V_{J+1}$, which is defined by

$$(\mathscr{I}_J v)(x) = \sum_{e \in \mathscr{E}_h} v(m_e)\psi_e(x),$$

where ψ_e is the CR basis on the edge/face $e \in \mathscr{E}_h$ and m_e is the barycenter of e. Therefore, the prolongation matrix has the same sparsity pattern as the edge-to-vertex (in 2D), or face-to-vertex (in 3D) connectivity, and each nonzero entry in this matrix equals the constant $1/d$ where d is the space dimension. The restriction matrix is simply the transpose of the prolongation matrix.

The efficiency and robustness of this preconditioner can be analyzed in terms of the *effective condition number* (cf. [12]) defined as follows:

Definition 1. *Let V be a real N dimensional Hilbert space, and $S : V \to V$ be a symmetric positive definition operator with eigenvalues $0 < \lambda_1 \leq \cdots \leq \lambda_N$. The m-th* effective condition *number of S is defined by*

$$\mathscr{K}_m(S) := \lambda_N(S)/\lambda_{m+1}(S).$$

Note that the standard condition number $\mathscr{K}(BA)$ of the preconditioned system BA will be large due to the large jump in the coefficient κ. However, there might be only a small (fixed) number of small eigenvalues of BA, which cause the large condition number; and the other eigenvalues are bounded nearly uniformly. In particular, we have the following main result:

Theorem 1. *Let B be the multigrid V-cycle preconditioner defined in Algorithm 3. Then there exists a fixed integer $m_0 < M$, depending only on the distribution of the coefficient κ, such that*

$$\mathscr{K}_{m_0}(BA) \leq C^2 |\log h|^2 = C^2 J^2 ,$$

where the constant $C > 0$ is independent of the coefficients and mesh size.

The analysis is based on the subspace correction framework [10], but some technical tools developed in [2] are needed to deal with nonconformity of the finite element spaces. Due to space restriction, a detailed analysis will be reported somewhere else.

Thanks to Theorem 1 and a standard PCG convergence result (cf. [1, Sect. 13.2]), the PCG algorithm with the multigrid V-cycle preconditioner defined in Algorithm 3 has the following convergence estimate:

$$|u - u_i|_{h,\kappa} \leq 2(\mathscr{K}(BA) - 1)^{m_0} \left(\frac{CJ-1}{CJ+1}\right)^{i-m_0} |u - u_0|_{h,\kappa} ,$$

where u_0 is the initial guess, and u_i is the solution of i-th PCG iteration. Although the condition number $\mathscr{K}(BA)$ might be large, the convergence rate of the PCG algorithm is asymptotically dominated by $\frac{CJ-1}{CJ+1}$, which is determined by the effective condition number $\mathscr{K}_{m_0}(BA)$. Moreover, this bound of asymptotic convergence rate convergence is independent of the coefficient κ, but depends on the mesh size logarithmically.

4 Numerical Results

In this section, we present several numerical tests in 2D and 3D which verify the result in Theorem 1 on the performance of the multigrid V-cycle preconditioner described in the previous sections. The numerical tests show that the effective condition numbers of the preconditioned linear systems (with V-cycle preconditioner) are nearly uniformly bounded.

4.1 A 2D Example

As a first model problem, we consider Eq. (1) in the square $\Omega = (-1,1)^2$ with coefficient such that, $\kappa(x) = 1$ for $x \in \Omega_1 = (-0.5,0)^2 \cup (0,0.5)^2$, and $\kappa(x) = \varepsilon$ for x in the remaining subdomain, $x \in \Omega \setminus \Omega_1$ (see Fig. 1). By decreasing the value of ε we increase the contrast in the PDE coefficients.

Our initial triangulation on level 0 has mesh size $h_0 = 2^{-1}$ and resolves the interfaces where the coefficients have discontinuities. Then on each level, we uniformly refine the mesh by subdividing each element into four congruent children. In this example, we use 1 forward/backward Gauss-Seidel iteration as pre/post smoother in the multigrid preconditioner, and the stopping criteria of the PCG algorithm is $\|r_k\|/\|r_0\| < 10^{-7}$ where r_k is the residual at k-th iteration.

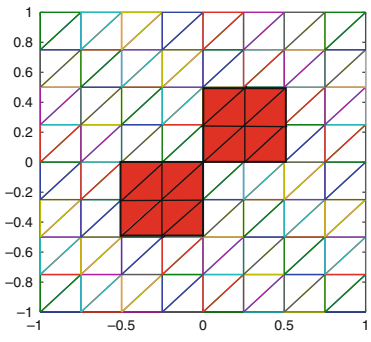


Fig. 1. 2D computational domain

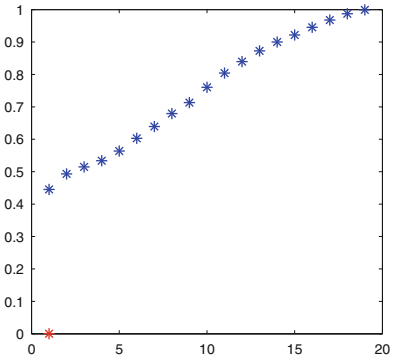


Fig. 2. Eigenvalue distribution of BA

Figure 2 shows the eigenvalue distribution of the multigrid V-cycle preconditioned system BA when $h = 2^{-5}$ (level = 4) and $\varepsilon = 10^{-5}$. As we can see from this figure, there is only one small eigenvalue that deteriorates with respect to the jump in the coefficient and the mesh size.

Table 4.1 shows the estimated condition number $\mathscr{K}$ and the effective condition number $\mathscr{K}_1$ of BA. It can be observed that the condition number $\mathscr{K}$ increases rapidly with respect to the increase of the jump in the coefficients and the number of degrees of freedom. On the other hand, the number of PCG iterations increases only a

small amount, and the corresponding effective condition number is nearly uniformly bounded, as predicted by Theorem 1.

ε	levels	0	1	2	3	4
1	$\mathscr{K}$	1.65 (8)	1.83 (10)	1.9 (10)	1.9 (10)	1.89 (10)
	$\mathscr{K}_1$	1.44	1.78	1.77	1.78	1.76
10^{-1}	$\mathscr{K}$	3.78 (10)	3.69 (11)	3.76 (12)	3.79 (12)	3.88 (12)
	$\mathscr{K}_1$	1.89	1.87	1.93	1.92	1.95
10^{-2}	$\mathscr{K}$	23.4 (12)	23.6 (13)	24.6 (13)	25.1 (14)	26 (15)
	$\mathscr{K}_1$	2.15	1.96	1.99	1.97	2.24
10^{-3}	$\mathscr{K}$	218 (13)	223 (14)	232 (15)	238 (16)	246 (16)
	$\mathscr{K}_1$	2.19	1.98	2	1.98	2.29
10^{-4}	$\mathscr{K}$	2.17e+3 (14)	2.21e+3 (15)	2.31e+3 (16)	2.37e+3 (18)	2.45e+3 (18)
	$\mathscr{K}_1$	2.2	1.98	2	1.98	2.3
10^{-5}	$\mathscr{K}$	2.17e+4 (15)	2.21e+4 (16)	2.31e+4 (17)	2.37e+4 (19)	2.76e+4 (19)
	$\mathscr{K}_1$	2.2	1.98	2	1.98	2.64

Table 1. Estimated condition number $\mathscr{K}$ (number of PCG iterations) and the effective condition number $\mathscr{K}_1$

4.2 A 3D Example

In this second example, we consider the model problem (1) in the open unit cube in 3D with a similar setting for the coefficient. We set $\kappa(x) = 1$ for $x \in \Omega_1 = (0.25, 0.5)^3$ or $x \in \Omega_2 = (0.5, 0.75)^3$, and $\kappa(x) = \varepsilon$ for the remaining subdomain (that is, for $x \in \Omega \setminus (\Omega_1 \cup \Omega_2)$). The domain Ω and the subdomains just described are shown in Fig. 3. The coarsest partition has mesh size $h_0 = 2^{-2}$, and it is set in a way so that it resolves the interfaces where the coefficient has jumps.

To test the effects of the smoother, in this example we use 5 forward/backward Gauss-Seidel as smoother in the multigrid preconditioner. In order to test more severe jumps in the coefficients, we set the stopping criteria $\|r_k\|/\|r_0\| < 10^{-12}$ for the PCG algorithm in this experiment.

Figure 4 shows the eigenvalue distribution of the multigrid V-cycle preconditioned system BA when $h = 2^{-5}$ (level = 3) and $\varepsilon = 10^{-5}$. As before, this figure shows that there is only one small eigenvalue that even deteriorates with respect to the jump in the coefficients and the mesh size.

Table 2 shows the estimated condition number $\mathscr{K}$ (with the number of PCG iterations), and the effective condition number $\mathscr{K}_1$. As is easily seen from the results in this table, the condition number $\mathscr{K}$ increases when ε decreases, i.e. the condition number grows when the jump in the coefficients becomes larger. On the other hand, the results in Table 2 show that the effective condition number $\mathscr{K}_1$ remains nearly uniformly bounded with respect to the mesh size and it is robust with respect to the jump in the coefficient, thus confirming the result stated in Theorem 1: a PCG with

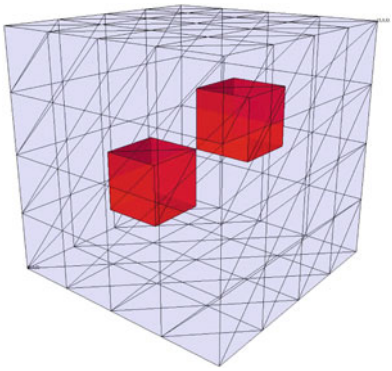


Fig. 3. 3D computational domain

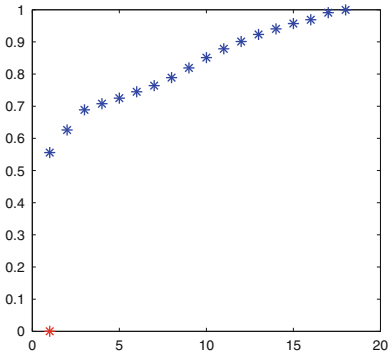


Fig. 4. Eigenvalue distribution of BA

ε	levels	0	1	2	3
1	$\mathscr{K}$	1.19 (8)	1.34 (11)	1.37 (11)	1.36 (11)
	$\mathscr{K}_1$	1.16	1.26	1.31	1.29
10^{-1}	$\mathscr{K}$	2.3 (10)	1.94(13)	1.75 (13)	1.67 (14)
	$\mathscr{K}_1$	1.60	1.56	1.45	1.43
10^{-3}	$\mathscr{K}$	86.01 (11)	63.07 (16)	52.67 (17)	48.19(17)
	$\mathscr{K}_1$	2.4	2.12	1.89	1.78
10^{-5}	$\mathscr{K}$	8.39+3 (13)	6.15e+3 (18)	5.13e+3 (19)	4.70e+3(19)
	$\mathscr{K}_1$	2.44	2.14	1.91	1.80
10^{-7}	$\mathscr{K}$	8.39+5 (14)	6.15e+5 (21)	5.13e+5 (23)	4.70e+5(21)
	$\mathscr{K}_1$	2.45	2.14	1.91	1.80

Table 2. Estimated condition number $\mathscr{K}$ (number of PCG iterations) and effective condition number $\mathscr{K}_1$.

multigrid V-cycle preconditioner provides a robust, nearly optimal solver for the CR approximation to (3).

Acknowledgments First author has been supported by MINECO grant MTM2011-27739-C04-04 and 2009-SGR-345 from AGAUR-Generalitat de Catalunya. The work of the second and third authors was supported in part by NSF/DMS Awards 0715146 and 0915220, and by DOD/DTRA Award HDTRA-09-1-0036. The work of the fourth author was supported in part by the NSF/DMS Award 0810982.

Bibliography

[1] O. Axelsson. *Iterative solution methods*. Cambridge University Press, Cambridge, 1994.

[2] B. Ayuso de Dios, M. Holst, Y. Zhu, and L. Zikatanov. Multilevel Preconditioners for Discontinuous Galerkin Approximations of Elliptic Problems with

Jump Coefficients. *Arxiv preprint arXiv:1012.1287*, to appear in Mathematics of Computation, 2010.

[3] J. H. Bramble. *Multigrid Methods*, volume 294 of *Pitman Research Notes in Mathematical Sciences*. Longman Scientific & Technical, Essex, England, 1993.

[4] W. L. Briggs, V. E. Henson, and S. F. McCormick. *A multigrid tutorial*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, second edition, 2000.

[5] M. Dryja, M. V. Sarkis, and Olof B. Widlund. Multilevel Schwarz methods for elliptic problems with discontinuous coefficients in three dimensions. *Numerische Mathematik*, 72(3):313–348, 1996.

[6] P. Oswald. Preconditioners for nonconforming discretizations. *Mathematics of Computation*, 65(215):923–941, 1996.

[7] M. Sarkis. Multilevel methods for P_1 nonconforming finite elements and discontinuous coefficients in three dimensions. In *Domain decomposition methods in scientific and engineering computing (University Park, PA, 1993)*, volume 180 of *Contemp. Math.*, pages 119–124. Amer. Math. Soc., Providence, RI, 1994.

[8] M. V. Sarkis. *Schwarz Preconditioners for Elliptic Problems with Discontinuous Coefficients Using Conforming and Non-Conforming Elements*. PhD thesis, Courant Institute of Mathematical Science of New York University, 1994.

[9] J. Xu. *Theory of Multilevel Methods*. PhD thesis, Cornell University, 1989.

[10] J. Xu. Iterative methods by space decomposition and subspace correction. *SIAM Review*, 34:581–613, 1992.

[11] J. Xu. The auxiliary space method and optimal multigrid preconditioning techniques for unstructured meshes. *Computing*, 56:215–235, 1996.

[12] J. Xu and Y. Zhu. Uniform convergent multigrid methods for elliptic problems with strongly discontinuous coefficients. *Mathematical Models and Methods in Applied Science*, 18(1):77 –105, 2008.

[13] Y. Zhu. Analysis of a multigrid preconditioner for Crouzeix-Raviart discretization of elliptic PDE with jump coefficient. *Arxiv preprint arXiv:1110.5159*, Numerical Linear Algebra with Applications, DOI: 10.1002/nla.1856, 2012.

Domain Decomposition Methods of Stochastic PDEs

Waad Subber and Abhijit Sarkar

Department of Civil and Environmental Engineering, Carleton University, Ottawa, Ontario K1S5B6, Canada waad_subber@carleton.ca, abhijit_sarkar@carleton.ca

1 Introduction

In conjunction with modern high performance computing systems, domain decomposition algorithms permit simulation of PDEs with extremely high resolution numerical models. Such computational models substantially reduce discretization errors. In realistic simulation of certain physical systems, it is however necessary to consider the heterogeneities of the model parameters. Whenever sufficient statistical information is available, such heterogeneities can be modeled by stochastic processes (e.g. [2]). For uncertainty propagation, the traditional Monte Carlo simulation may be impractical for these high resolution models. As an alternative, a domain decomposition algorithm for stochastic PDEs (SPDEs) is proposed [4] using the spectral stochastic finite element method (SSFEM). The SSFEM discretization leads to a linear system with a block sparsity structure, and the size of the resulting system grows rapidly with the spatial mesh resolution and the order of the stochastic dimension [2]. The solution of this large-scale system constitutes a computationally challenging task and therefore efficient solvers are required. Extending the formulation in [4], the iterative substructuring based non-overlapping domain decomposition methods are proposed to solve the large-scale linear system arising in the SSFEM. The methodology is based on domain decomposition in the geometric space and a functional decomposition in the stochastic space [4]. Firstly, we describe a primal version of iterative substructuring methods of SPDEs. The method offers a straightforward approach to formulate a two-level scalable preconditioner. In the proposed preconditioner, the continuity of the solution field is strictly enforced on the corner nodes of the interface boundary, but weakly satisfied over the remaining interface nodes. This approach naturally leads to a coarse grid connecting the subdomains globally and provides a mechanism to propagate information across the subdomains which makes the algorithm scalable. The proposed preconditioner may be viewed as an extension of BDDC [3] for SPDEs. Secondly, a dual-primal iterative substructuring method is introduced for SPDEs. In this approach, the continuity condition on the corner nodes is strictly satisfied and Lagrange multipliers are used to weakly enforce the continu-

R. Bank et al. (eds.), *Domain Decomposition Methods in Science and Engineering XX*, Lecture Notes in Computational Science and Engineering 91,
DOI 10.1007/978-3-642-35275-1_21,

ity on the remaining nodes of the interface boundary. This method may be construed to be an extension of FETI-DP [1] for SPDEs.

2 Uncertainty Representation by Stochastic Processes

We briefly review the theories of stochastic processes, relevant to subsequent theoretical developments, by closely following [2, 4–6]. Assuming the input data (containing sufficient statistical information) permits a representation of the model parameters as stochastic processes that span the Hilbert space $\mathscr{H}_G$. Using Karhunen-Loeve expansion (KLE), a set of basis functions $\{\xi_i(\theta)\}$ for the Hilbert space $\mathscr{H}_G$ is identified. The KLE of a stochastic process $\alpha(\mathbf{x},\theta)$ is based on the spectral expansion of its covariance function $C_{\alpha\alpha}(\mathbf{x},\mathbf{y})$, and takes the following form [2]

$$\alpha(\mathbf{x},\theta) = \bar{\alpha}(\mathbf{x}) + \sum_{i=1}^{\infty} \sqrt{\lambda_i}\xi_i(\theta)\phi_i(\mathbf{x}), \tag{1}$$

where $\bar{\alpha}(\mathbf{x})$ is the mean of the stochastic process, $\{\xi_i(\theta)\}$ is a set of uncorrelated random variables and $\{\lambda_i,\phi_i(\mathbf{x})\}$ are the eigenpairs of the covariance function, obtained from the following integral equation

$$\int_{\Omega} C_{\alpha\alpha}(\mathbf{x},\mathbf{y})\phi_i(\mathbf{y})d\mathbf{y} = \lambda_i\phi_i(\mathbf{x}). \tag{2}$$

For a smooth stochastic process, only a finite number of KLE basis is sufficient to represent the stochastic process. Given the covariance function of the solution is not known a priori, the KLE cannot be used to represent solution process. Assuming the solution process $u(\mathbf{x},\theta)$ belong to the Hilbert space $\mathscr{H}_L$, a generic basis of this space can be identified using the Polynomial Chaos (PC) [2]. Consequently, the solution process can be approximated as

$$u(\mathbf{x},\theta) = \sum_{j=0}^{N} \Psi_j(\theta)u_j(\mathbf{x}), \tag{3}$$

where the polynomials $\Psi_j(\theta)$ are orthogonal in the statistical sense, meaning $\langle\Psi_j,\Psi_k\rangle = \langle\Psi_j^2\rangle\delta_{jk}$ where $\langle\cdot\rangle$ denotes the expectation operator and δ_{jk} is the Kronecker delta, and $u_j(\mathbf{x})$ are the PC coefficients to be determined by Galerkin projection.

3 Review of Schur Complement Based Domain Decomposition Method of SPDEs

A review of the domain decomposition method for SPDEs based on [4–6] is provided in this section. For an elliptic SPDE defined on a domain Ω with a prescribed

boundary condition on $\partial\Omega$, the finite element discretization leads to the following linear system

$$\mathbf{A}(\theta)\mathbf{u}(\theta) = \mathbf{f}, \tag{4}$$

where $\mathbf{A}(\theta)$ is the random stiffness matrix, $\mathbf{u}(\theta)$ is the stochastic response and $\mathbf{f}$ is the applied force. The physical domain Ω is split into n_s non-overlapping subdomains $\{\Omega_s\}_{s=1}^{n_s}$. For a typical subdomain Ω_s the nodal vector $\mathbf{u}^s(\theta)$ is partitioned into interior $\mathbf{u}_I^s(\theta)$ and interface $\mathbf{u}_\Gamma^s(\theta)$ unknowns. This decomposition leads to the following subdomain equilibrium equation

$$\begin{bmatrix} \mathbf{A}_{II}^s(\theta) & \mathbf{A}_{I\Gamma}^s(\theta) \\ \mathbf{A}_{\Gamma I}^s(\theta) & \mathbf{A}_{\Gamma\Gamma}^s(\theta) \end{bmatrix} \begin{Bmatrix} \mathbf{u}_I^s(\theta) \\ \mathbf{u}_\Gamma^s(\theta) \end{Bmatrix} = \begin{Bmatrix} \mathbf{f}_I^s \\ \mathbf{f}_\Gamma^s \end{Bmatrix}. \tag{5}$$

Enforcing the transmission conditions and expanding the solution vector by the PCE (as in Eq. (3)) and then performing Galerkin projection, we obtain the following block linear systems of equations [4–6]:

$$\begin{aligned} \langle \sum_{i=0}^{L} \Psi_i(\theta) & \begin{bmatrix} \mathbf{A}_{II,i}^1 & \dots & 0 & \mathbf{A}_{I\Gamma,i}^1 \mathbf{R}_1 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & \mathbf{A}_{II,i}^{n_s} & \mathbf{A}_{I\Gamma,i}^{n_s} \mathbf{R}_{n_s} \\ \mathbf{R}_1^T \mathbf{A}_{\Gamma I,i}^1 & \dots & \mathbf{R}_{n_s}^T \mathbf{A}_{\Gamma I,i}^{n_s} & \sum_{s=1}^{n_s} \mathbf{R}_s^T \mathbf{A}_{\Gamma\Gamma,i}^s \mathbf{R}_s \end{bmatrix} \sum_{j=0}^{N} \Psi_j(\theta) \begin{Bmatrix} \mathbf{u}_{I,j}^1 \\ \vdots \\ \mathbf{u}_{I,j}^{n_s} \\ \mathbf{u}_{\Gamma,j} \end{Bmatrix} \Psi_k(\theta) \rangle \\ & = \langle \begin{Bmatrix} \mathbf{f}_I^1 \\ \vdots \\ \mathbf{f}_I^{n_s} \\ \sum_{s=1}^{n_s} \mathbf{R}_s^T \mathbf{f}_\Gamma^s \end{Bmatrix} \Psi_k(\theta) \rangle, \quad k = 0, \dots, N. \end{aligned} \tag{6}$$

where the restriction operator $\mathbf{R}_s$ maps the global interface vector $\mathbf{u}_\Gamma(\theta)$ to the local interface unknown $\mathbf{u}_\Gamma^s(\theta)$ as $\mathbf{u}_\Gamma^s(\theta) = \mathbf{R}_s \mathbf{u}_\Gamma(\theta)$. Compactly, Eq. (6) can be expressed as

$$\begin{bmatrix} \mathscr{A}_{II}^1 & \dots & 0 & \mathscr{A}_{I\Gamma}^1 \mathscr{R}_1 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & \mathscr{A}_{II}^{n_s} & \mathscr{A}_{I\Gamma}^{n_s} \mathscr{R}_{n_s} \\ \mathscr{R}_1^T \mathscr{A}_{\Gamma I}^1 & \dots & \mathscr{R}_{n_s}^T \mathscr{A}_{\Gamma I}^{n_s} & \sum_{s=1}^{n_s} \mathscr{R}_s^T \mathscr{A}_{\Gamma\Gamma}^s \mathscr{R}_s \end{bmatrix} \begin{Bmatrix} \mathscr{U}_I^1 \\ \vdots \\ \mathscr{U}_I^{n_s} \\ \mathscr{U}_\Gamma \end{Bmatrix} = \begin{Bmatrix} \mathscr{F}_I^1 \\ \vdots \\ \mathscr{F}_I^{n_s} \\ \sum_{s=1}^{n_s} \mathscr{R}_s^T \mathscr{F}_\Gamma^s \end{Bmatrix}, \tag{7}$$

where $[\mathscr{A}_{\alpha\beta}^s]_{jk} = \sum_{i=0}^{L} \langle \Psi_i \Psi_j \Psi_k \rangle \mathbf{A}_{\alpha\beta,i}^s$, $\mathscr{F}_{\alpha,k}^s = \langle \Psi_k \mathbf{f}_\alpha^s \rangle$, $\mathscr{U}_I^m = (\mathbf{u}_{I,0}^m, \dots, \mathbf{u}_{I,N}^m)^T$ and $\mathscr{R}_s = blockdiag(\mathbf{R}_s^0, \dots, \mathbf{R}_s^N)$. The subscripts α and β represent the index I and Γ. Performing Gaussian elimination in Eq. (7), we obtain the global *extended* Schur complement system as

$$\mathscr{S}\, \mathscr{U}_\Gamma = \mathscr{G}_\Gamma, \tag{8}$$

where $\mathscr{S} = \sum_{s=1}^{n_s} \mathscr{R}_s^T [\mathscr{A}_{\Gamma\Gamma}^s - \mathscr{A}_{\Gamma I}^s (\mathscr{A}_{II}^s)^{-1} \mathscr{A}_{I\Gamma}^s] \mathscr{R}_s$, $\mathscr{G}_\Gamma = \sum_{s=1}^{n_s} \mathscr{R}_s^T [\mathscr{F}_\Gamma^s - \mathscr{A}_{\Gamma:I}^s (\mathscr{A}_{II}^s)^{-1} \mathscr{F}_I^s]$.

4 Primal Iterative Substructuring Method of SPDEs

In this section, a two-level domain decomposition method is formulated in the context of SPDEs. The subdomain nodal vector, namely the primal variable, is partitioned into interior, remaining interface and corner nodes as schematically shown in Fig. 1 [3]. Using PCE to represent the random coefficients of the system parameters and performing Galerkin projection, lead to the following coupled deterministic system

$$\begin{bmatrix} \mathcal{A}_{ii}^s & \mathcal{A}_{ir}^s & \mathcal{A}_{ic}^s \\ \mathcal{A}_{ri}^s & \mathcal{A}_{rr}^s & \mathcal{A}_{rc}^s \\ \mathcal{A}_{ci}^s & \mathcal{A}_{cr}^s & \mathcal{A}_{cc}^s \end{bmatrix} \begin{Bmatrix} \mathcal{U}_i^s \\ \mathcal{U}_r^s \\ \mathcal{U}_c^s \end{Bmatrix} = \begin{Bmatrix} \mathcal{F}_i^s \\ \mathcal{F}_r^s \\ \mathcal{F}_c^s \end{Bmatrix}. \tag{9}$$

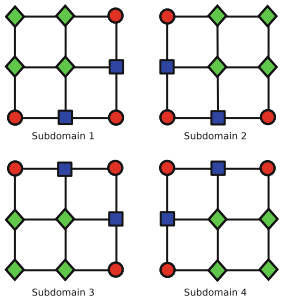


Fig. 1. Partitioning domain nodes into: interior (♦), remaining (■) and corner (●)

Enforcing the transmission conditions along the boundary interfaces, the subdomain equilibrium equation can be written as

$$\begin{bmatrix} \mathcal{A}_{ii}^s & \mathcal{A}_{ir}^s\mathcal{B}_r^s & \mathcal{A}_{ic}^s\mathcal{B}_c^s \\ \sum_{s=1}^{n_s}\mathcal{B}_r^{sT}\mathcal{A}_{ri}^s & \sum_{s=1}^{n_s}\mathcal{B}_r^{sT}\mathcal{A}_{rr}^s\mathcal{B}_r^s & \sum_{s=1}^{n_s}\mathcal{B}_r^{sT}\mathcal{A}_{rc}^s\mathcal{B}_c^s \\ \sum_{s=1}^{n_s}\mathcal{B}_c^{sT}\mathcal{A}_{ci}^s & \sum_{s=1}^{n_s}\mathcal{B}_c^{sT}\mathcal{A}_{cr}^s\mathcal{B}_r^s & \sum_{s=1}^{n_s}\mathcal{B}_c^{sT}\mathcal{A}_{cc}^s\mathcal{B}_c^s \end{bmatrix} \begin{Bmatrix} \mathcal{U}_i^s \\ \mathcal{U}_r \\ \mathcal{U}_c \end{Bmatrix} = \begin{Bmatrix} \mathcal{F}_i^s \\ \sum_{s=1}^{n_s}\mathcal{B}_r^{sT}\mathcal{F}_r^s \\ \sum_{s=1}^{n_s}\mathcal{B}_c^{sT}\mathcal{F}_c^s \end{Bmatrix}, \tag{10}$$

where $\mathcal{B}_r^s$ and $\mathcal{B}_c^s$ are Boolean rectangular matrices that extract the subdomain remaining interface and corner degrees of freedom from the corresponding global vectors $\mathcal{U}_r$ and $\mathcal{U}_c$ as $\mathcal{U}_r^s = \mathcal{B}_r^s\mathcal{U}_r$ and $\mathcal{U}_c^s = \mathcal{B}_c^s\mathcal{U}_c$. Eliminating $\mathcal{U}_i^s$ from Eq. (10), we obtain

$$\begin{bmatrix} \sum_{s=1}^{n_s}\mathcal{B}_r^{sT}\mathcal{S}_{rr}^s\mathcal{B}_r^s & \sum_{s=1}^{n_s}\mathcal{B}_r^{sT}\mathcal{S}_{rc}^s\mathcal{B}_c^s \\ \sum_{s=1}^{n_s}\mathcal{B}_c^{sT}\mathcal{S}_{cr}^s\mathcal{B}_r^s & \sum_{s=1}^{n_s}\mathcal{B}_c^{sT}\mathcal{S}_{cc}^s\mathcal{B}_c^s \end{bmatrix} \begin{Bmatrix} \mathcal{U}_r \\ \mathcal{U}_c \end{Bmatrix} = \begin{Bmatrix} \sum_{s=1}^{n_s}\mathcal{B}_r^{sT}\mathcal{G}_r^s \\ \sum_{s=1}^{n_s}\mathcal{B}_c^{sT}\mathcal{G}_c^s \end{Bmatrix}, \tag{11}$$

where $\mathcal{S}^s_{\alpha\beta} = \mathcal{A}^s_{\alpha\beta} - \mathcal{A}^s_{\alpha i}[\mathcal{A}^s_{ii}]^{-1}\mathcal{A}^s_{i\beta}$ and $\mathcal{G}^s_\alpha = \mathcal{F}^s_\alpha - \mathcal{A}^s_{\alpha i}[\mathcal{A}^s_{ii}]^{-1}\mathcal{F}^s_i$. Eliminating $\mathcal{U}_c$ from Eq. (11) leads to the following symmetric positive definite *reduced interface problem*

$$(F_{rr} - F_{rc}[F_{cc}]^{-1}F_{cr})\mathcal{U}_r = d_r - F_{rc}[F_{cc}]^{-1}d_c, \tag{12}$$

where $F_{\alpha\beta} = \sum_{s=1}^{n_s} \mathcal{B}^{s\,T}_\alpha \mathcal{S}^s_{\alpha\beta}\mathcal{B}^s_\beta$ and $d_\alpha = \sum_{s=1}^{n_s} \mathcal{B}^{s\,T}_\alpha \mathcal{G}^s_\alpha$.

4.1 Two-Level Primal Preconditioner

The Preconditioned Conjugate Gradient Method (PCGM) can be used to solve the reduced interface problem in Eq. (12). At each iteration of the PCGM, the continuity of the solution field is enforced strictly on the corner nodes, but weakly satisfied on the remaining interface nodes. Consequently we obtain the following partially assembled Schur complement system:

$$\begin{bmatrix} \mathcal{S}^s_{rr} & \mathcal{S}^s_{rc}\mathcal{B}^s_c \\ \sum_{s=1}^{n_s} \mathcal{B}^{s\,T}_c \mathcal{S}^s_{cr}\mathcal{B}^s_r & \sum_{s=1}^{n_s} \mathcal{B}^{s\,T}_c \mathcal{S}^s_{cc}\mathcal{B}^s_c \end{bmatrix} \begin{Bmatrix} \mathcal{U}^s_r \\ \mathcal{U}_c \end{Bmatrix} = \begin{Bmatrix} \mathcal{F}^s_r \\ 0 \end{Bmatrix}, \tag{13}$$

where $\mathcal{F}^s_r = \mathcal{D}^s_r \mathcal{B}^s_r \mathbf{r}_j$, and $\mathbf{r}_j$ is the residual of the jth iteration of PCGM, and $\mathcal{D}^s_r$ is a block diagonal weighting matrix which satisfies $\sum_{s=1}^{n_s} \mathcal{B}^{s\,T}_r \mathcal{D}^s_r \mathcal{B}^s_r = \mathbf{I}$. Next, $\mathcal{U}^s_r$ can be eliminated from Eq. (13) leading to the following coarse problem

$$\widetilde{F}_{cc}\mathcal{U}_c = \widetilde{d}_c, \tag{14}$$

where $\widetilde{F}_{cc} = \sum_{s=1}^{n_s} \mathcal{B}^{s\,T}_c (\mathcal{S}^s_{cc} - \mathcal{S}^s_{cr}[\mathcal{S}^s_{rr}]^{-1}\mathcal{S}^s_{rc})\mathcal{B}^s_c$ and $\widetilde{d}_c = -\sum_{s=1}^{n_s} \mathcal{B}^{s\,T}_c \mathcal{S}^s_{cr}[\mathcal{S}^s_{rr}]^{-1}\mathcal{F}^s_r$. The two-level preconditioner can be expressed as

$$\mathcal{M}^{-1} = \sum_{s=1}^{n_s} \mathcal{B}^{s\,T}_r \mathcal{D}^s_r [\mathcal{S}^s_{rr}]^{-1} \mathcal{D}^s_r \mathcal{B}^s_r + R_0^T [\widetilde{F}_{cc}]^{-1} R_0, \tag{15}$$

where $R_0 = \sum_{s=1}^{n_s} \mathcal{B}^{s\,T}_c \mathcal{S}^s_{cr}[\mathcal{S}^s_{rr}]^{-1}\mathcal{D}^s_r \mathcal{B}^s_r$.

5 Dual-Primal Iterative Substructuring of SPDEs

In the dual-primal method [1], the continuity condition on the corner nodes is enforced strictly while Lagrange multipliers are used to weakly enforce the continuity on the remaining interface. Partial assembly of the corner node unknowns leads to the following system

$$\begin{bmatrix} \mathscr{A}_{ii}^s & \mathscr{A}_{ir}^s & \mathscr{A}_{ic}^s\mathscr{B}_c^s & 0 \\ \mathscr{A}_{ri}^s & \mathscr{A}_{rr}^s & \mathscr{A}_{rc}^s\mathscr{B}_c^s & \mathscr{B}_r^{sT} \\ \sum_{s=1}^{n_s}\mathscr{B}_c^{sT}\mathscr{A}_{ci}^s & \sum_{s=1}^{n_s}\mathscr{B}_c^{sT}\mathscr{A}_{cr}^s & \sum_{s=1}^{n_s}\mathscr{B}_c^{sT}\mathscr{A}_{cc}^s\mathscr{B}_c^s & 0 \\ 0 & \sum_{s=1}^{n_s}\mathscr{B}_r^s & 0 & 0 \end{bmatrix} \begin{Bmatrix} \mathscr{U}_i^s \\ \mathscr{U}_r^s \\ \mathscr{U}_c \\ \Lambda \end{Bmatrix} = \begin{Bmatrix} \mathscr{F}_i^s \\ \mathscr{F}_r^s \\ \sum_{s=1}^{n_s}\mathscr{B}_c^{sT}\mathscr{F}_c^s \\ 0 \end{Bmatrix}, \tag{16}$$

where $\sum_{s=1}^{n_s}\mathscr{B}_r^s\mathscr{U}_r^s = 0$ and $\Lambda^T = \{\boldsymbol{\lambda}_0, \cdots, \boldsymbol{\lambda}_N\}$. The matrix $\mathscr{B}_r^s$ is a block diagonal signed Boolean continuity operator and $\boldsymbol{\lambda}_j$ is the nodal force vector required to satisfy continuity on the remaining interface nodes. Eliminating $\mathscr{U}_i^s$ and $\mathscr{U}_r^s$ from Eq. (16) leads to the following interface problem

$$\begin{bmatrix} \bar{F}_{cc} & -\bar{F}_{cr} \\ \bar{F}_{rc} & \bar{F}_{rr} \end{bmatrix} \begin{Bmatrix} \mathscr{U}_c \\ \Lambda \end{Bmatrix} = \begin{Bmatrix} \bar{d}_c \\ \bar{d}_r \end{Bmatrix}, \tag{17}$$

where

$$\bar{F}_{cc} = \sum_{s=1}^{n_s}\mathscr{B}_c^{sT}(\mathscr{S}_{cc}^s - \mathscr{S}_{cr}^s[\mathscr{S}_{rr}^s]^{-1}\mathscr{S}_{rc}^s)\mathscr{B}_c^s, \quad \bar{F}_{cr} = \sum_{s=1}^{n_s}\mathscr{B}_c^{sT}\mathscr{S}_{cr}^s[\mathscr{S}_{rr}^s]^{-1}\mathscr{B}_r^{sT}$$

$$\bar{F}_{rc} = \sum_{s=1}^{n_s}\mathscr{B}_r^s[\mathscr{S}_{rr}^s]^{-1}\mathscr{S}_{rc}^s\mathscr{B}_c^s, \qquad \bar{F}_{rr} = \sum_{s=1}^{n_s}\mathscr{B}_r^s[\mathscr{S}_{rr}^s]^{-1}\mathscr{B}_r^{sT}$$

$$\bar{d}_c = \sum_{s=1}^{n_s}\mathscr{B}_c^{sT}(\mathscr{G}_c^s - \mathscr{S}_{cr}^s[\mathscr{S}_{rr}^s]^{-1}\mathscr{G}_r^s), \qquad \bar{d}_r = \sum_{s=1}^{n_s}\mathscr{B}_r^s[\mathscr{S}_{rr}^s]^{-1}\mathscr{G}_r^s$$

Solving for $\mathscr{U}_c$ from Eq. (17) gives the following coarse problem

$$\bar{F}_{cc}\mathscr{U}_c = (\bar{d}_c + \bar{F}_{cr}\Lambda) \tag{18}$$

Substituting $\mathscr{U}_c$ into Eq. (17) leads to the following symmetric positive definite Lagrange multiplier system

$$(\bar{F}_{rr} + \bar{F}_{rc}[\bar{F}_{cc}]^{-1}\bar{F}_{cr})\Lambda = \bar{d}_r - \bar{F}_{rc}[\bar{F}_{cc}]^{-1}\bar{d}_c. \tag{19}$$

The Lagrange multiplier system in Eq. (19) is solved using PCGM equipped with a Dirichlet precondtioner defined as $\bar{\mathscr{M}} = \sum_{s=1}^{n_s}\mathscr{B}_r^s\mathscr{D}_r^s\mathscr{S}_{rr}^s\mathscr{D}_r^s\mathscr{B}_r^{sT}$.

6 Numerical Results

For numerical illustrations, we consider the following elliptic SPDE

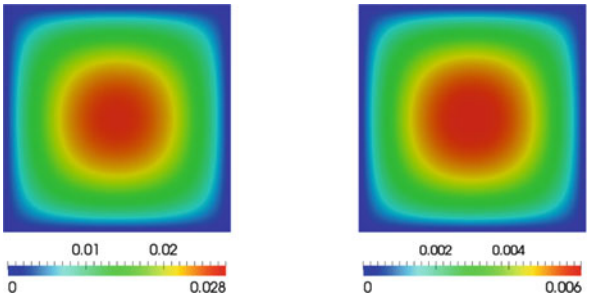
$$\nabla\cdot(\kappa(\mathbf{x},\theta)\nabla u(\mathbf{x},\theta)) = f(\mathbf{x}), \quad \mathbf{x}\in\Omega, \tag{20}$$

$$u(\mathbf{x},\theta) = 0, \quad \mathbf{x}\in\partial\Omega. \tag{21}$$

The coefficient $\kappa(\mathbf{x}, \theta)$ is modeled as a lognormal stochastic process, obtained from the underlying Gaussian process with an exponential covariance function given as

$$C_{\alpha\alpha}(\mathbf{x}, \mathbf{y}) = \sigma^2 \exp\left(-\frac{|x_1 - y_1|}{b_1} - \frac{|x_2 - y_2|}{b_2}\right). \tag{22}$$

The lognormal process is approximated using four-dimensional second order PC expansion ($L = 15$). Finite element discretization results in 375,444 elements and 186,925 nodes. The response is expressed using third order PCE ($N = 34$) leading to a linear system of order 6,542,375. The mean and standard deviation of the solution process are shown in Fig. 2. The PCGM iteration counts for the primal and dual-primal methods for fixed problem size in the spatial domain is reported in Table 1 for 1st, 2nd and 3rd order of PCE. The results suggest that the methods are numerically scalable with respect to number of subdomains. Table 2 shows the iteration counts of the methods when we fix spatial problem size per subdomain and increase the overall problem size by adding more subdomains. Again these results suggest that both the methods are numerically scalable with respect to fixed problem size per subdomain.



(a) (b)

Fig. 2. The mean and standard deviation of the solution field. **(a)** Mean. **(b)** Standard deviation

Table 1. Iteration counts for fixed problem size in geometric space

Subdomain	PP-DDM			DP-DDM		
	1*st*	2*nd*	3*rd*	1*st*	2*nd*	3*rd*
8	11	12	12	9	9	9
16	12	13	13	10	10	10
32	14	14	14	11	11	11
64	13	14	14	10	10	10
128	14	14	14	10	10	10
256	14	14	14	10	10	10

Table 2. Iteration counts for fixed problem size per subdomain in geometric space

Subdomain	PP-DDM			DP-DDM		
	1*st*	2*nd*	3*rd*	1*st*	2*nd*	3*rd*
8	9	9	9	8	8	8
16	12	12	12	10	10	10
32	12	13	13	10	10	10
64	13	14	14	10	10	10
128	14	14	14	10	10	10
256	15	15	15	11	11	11

7 Conclusion

Primal and dual-primal domain decomposition methods are proposed to solve the large-scale linear system arising from the finite element discretization of SPDEs. The proposed techniques exploit a coarse grid in the geometric space which makes the methods numerically scalable with respect to fixed geometric problem size, fixed geometric size per subdomain and the order of PCE.

Bibliography

[1] C. Farhat, M. Lesoinne, and K. Pierson. A scalable dual-primal domain decomposition method. *Numerical Linear Algebra with Applications*, 7:687–714, 2000.

[2] R. Ghanem and P. Spanos. *Stochastic Finite Element: A Spectral Approach*. Springer-Verlag, New York, 1991.

[3] J. Mandel and C. Dohrmann. Convergence of a balancing domain decomposition by constraints and energy minimization. *Numerical Linear Algebra with Applications*, 10:639–659, 2003.

[4] A. Sarkar, N. Benabbou, and R. Ghanem. Domain decomposition of stochastic PDEs: Theoretical formulations. *IJNME*, 77:689–701, 2009.

[5] W. Subber and A. Sarkar. Domain decomposition of stochastic PDEs: A novel preconditioner and its parallel performance. In *HPCS*, volume 5976 of *Lecture Notes in Computer Science*, pages 251–268. Springer, 2010.

[6] W. Subber and A. Sarkar. Primal and dual-primal iterative substructuring methods of stochastic PDEs. *Journal of Physics: Conference Series*, 256(1), 2010.

Improving the Convergence of Schwarz Methods for Helmholtz Equation

Murthy N Guddati and Senganal Thirunavukkarasu

North Carolina State University, Raleigh, North Carolina, USA.
murthy.guddati@ncsu.edu

1 Introduction

Various domain decomposition methods have been proposed for the Helmholtz equation, with the Optimized Schwarz Method (OSM) being one of them (see e.g. [7] for a review of various domain decomposition methods, and [3] for the details of OSM). In this paper, we focus on OSM, which is based on the idea of using approximated half-space Dirichlet-to-Neumann (DtN) maps to improve the convergence of the Schwarz methods; current version of the OSM is based on polynomial approximation of the half-space DtN map. See [8] for a review of various approaches to approximating the half-space DtN map (more commonly referred to as Absorbing Boundary Conditions (ABCs)).

There are two approximations in the OSM that affect its convergence rate – the first being the approximation of the rest of the domain as unbounded and the second being the approximation of the half-space stiffness (square-root operator) as a polynomial. In contrast with the polynomial approximation used in OSM, we utilize the method of Perfectly Matched Discrete Layers (PMDL), which has close links to the well-known Perfectly Matched Layers (PML) (see [1]) and the rational approximation of the square-root operator. The resulting PMDL-Schwarz method is shown to converge faster than the second-order OSM. The rest of the paper contains a brief review of OSM and PMDL concepts, followed by an outline of the new PMDL-Schwarz method and illustration of its effectiveness with the help of convergence factor analysis and a numerical example.

Model Problem. We consider the governing equation,

$$-\frac{\partial^2 \hat{u}}{\partial x^2} - \frac{\partial^2 \hat{u}}{\partial y^2} - \omega^2 \hat{u} = \hat{f}\,, \quad (x,y) \in (-\infty,\infty) \times [0,L], \tag{1a}$$

$$\hat{u}(\cdot,0) = \hat{u}(\cdot,L) = 0\,. \tag{1b}$$

Applying Fourier Sine transform along the y direction, the above equation reduces to a 1-D form:

R. Bank et al. (eds.), *Domain Decomposition Methods in Science and Engineering XX*, Lecture Notes in Computational Science and Engineering 91,
DOI 10.1007/978-3-642-35275-1_22,

$$-\frac{\partial^2 u}{\partial x^2} - k^2 u = f\,, \quad x \in (-\infty, \infty), \tag{2}$$

where $k = \sqrt{\omega^2 - k_y^2}$, k_y is the wavenumber along y and u, f are the Fourier symbols corresponding to $\hat{u}$, $\hat{f}$ respectively. For simplicity, we shall use the above 1-D equation to discuss the main ideas in this paper, but note that the proposed method is applicable to more complex equations and geometries. Also, since the focus of this paper is to improve the treatment of the transmission condition at an interface, it is sufficient to consider the case of two subdomains. Thus the domain is decomposed into two subdomains: $\Omega_1 \equiv (-\infty, 0)$ and $\Omega_2 \equiv (0, \infty)$, with the interface at $x = 0$.

2 Optimized Schwarz Methods

Optimized Schwarz Method is a domain decomposition method that is a variant of the Schwarz Alternating Method (see e.g. [7]). In the Schwarz Alternating Method, the displacement and traction continuity across the artificial interface are enforced by applying a mixed boundary condition of the form $\mathscr{B}(\cdot) \equiv \partial(\cdot)/\partial \mathbf{n} + \Lambda(\cdot)$ where $\mathbf{n}$ is the normal vector at the interface and the operator Λ is a parameter of the method. The Schwarz iteration scheme for solving (2) is given by:

$$-\frac{\partial^2 u_1^{j+1}}{\partial x^2} - k^2 u_1^{j+1} = f_1\,, \quad x \in \Omega_1\,, \qquad -\frac{\partial^2 u_2^{j+1}}{\partial x^2} - k^2 u_2^{j+1} = f_2\,, \quad x \in \Omega_2\,, \tag{3a}$$

$$\mathscr{B}_1 u_1^{j+1} = \mathscr{B}_1 u_2^{j}, \quad x = 0\,, \qquad \mathscr{B}_2 u_2^{j+1} = \mathscr{B}_2 u_1^{j}, \quad x = 0\,, \tag{3b}$$

$$\mathscr{B}_1(\cdot) \equiv \frac{\partial(\cdot)}{\partial \mathbf{n}_1} + \Lambda_1(\cdot)\,, \qquad \mathscr{B}_2(\cdot) \equiv \frac{\partial(\cdot)}{\partial \mathbf{n}_2} + \Lambda_2(\cdot)\,, \tag{3c}$$

where the operators $\Lambda_{1,2}$ are the parameters of the iteration that determine the convergence rate. The problem now reduces to choosing the parameters that lead to optimal convergence of the iteration scheme. The parameters are commonly chosen to be scalars but they can be operators that are optimized for convergence [3]. The dependence of the convergence on the choice of parameters is better understood by looking at the convergence factor ρ, which is defined as

$$\left|\hat{e_i}^{j+1}\right| = \rho \left|\hat{e_i}^{j}\right|, \tag{4}$$

where $\hat{e}_i^j = |u - u_i^j|$ is the error in the solution in subdomain i at iteration j. Thus, after one cycle of iteration, the error in solution reduces by ρ and the iterative scheme converges to a solution as long as $\rho < 1$.

For the Schwarz method in (3), the convergence factor can be shown to be (see for e.g. [3])

$$\rho = \left| \left(\frac{\Lambda_1 - \mathscr{K}_2}{\Lambda_1 + \mathscr{K}_1} \right) \left(\frac{\Lambda_2 - \mathscr{K}_1}{\Lambda_2 + \mathscr{K}_2} \right) \right|, \tag{5}$$

where $\mathscr{K}_1$ and $\mathscr{K}_2$ are the DtN maps of the subdomains Ω_1 and Ω_2 respectively. It is clear from (5) that the iterative scheme does not converge (because $\rho = 1$) for

a pure Neumann ($\Lambda_i = 0$) or Dirichlet ($\Lambda_i = \infty$) interface condition. Also, if $\Lambda_1 = \mathscr{K}_2$ or $\Lambda_2 = \mathscr{K}_1$, then $\rho = 0$ and the Schwarz iterative scheme converges in two iterations, i.e., the parameters are optimal. However, DtN maps are known only for special cases and even then are usually non-local operators that are expensive to compute accurately. Thus we look for local approximations to these DtN maps that are accurate and computationally efficient.

Optimized Schwarz Methods [3] essentially approximate the DtN map of the subdomains by polynomial approximations of the DtN map of an unbounded domain, e.g. the second-order OSM makes the approximation

$$\mathscr{K}_1 = -\mathrm{i}\sqrt{\omega^2 - k_y^2} \approx p + q k_y^2, \tag{6}$$

where p, q are parameters that are found by minimizing the convergence factor over the entire range of allowed vertical wavenumbers k_y. Note that there are other variants of OSM based on zeroth-order approximation; in this paper, we focus on the best available OSM, namely the second-order OSM.

3 A Schwarz Method with Improved Convergence

It appears to us that OSM uses polynomial approximation for reasons of implementability. A better approximation would be to use higher order rational approximations, which have been investigated extensively in the context of Absorbing Boundary Conditions (ABCs); it is now possible to implement these resulting ABCs and can also be used in the context of Schwarz methods. In this paper, we propose the use of a rational approximation in a recent ABC called Perfectly Matched Discrete Layers (formerly known as Continued Fraction ABCs – see [4]) instead of the polynomial approximation in (6).

The rational approximation corresponding to PMDL is given by:

$$\mathscr{K}_1 = -\mathrm{i}\sqrt{\omega^2 - k_y^2} \approx \mathscr{S}_n^{pmdl}, \tag{7}$$

where

$$\mathscr{S}_n^{pmdl} = p_n - \cfrac{q_n^2}{p_n + \left(p_{n-1} + \cfrac{q_{n-1}^2}{p_{n-1} + \left(p_{n-2} - \cfrac{q_{n-2}^2}{p_{n-2} + (\ldots)} \right)} \right)}, \tag{8}$$

$$\left.\begin{aligned} p_i &= \frac{1}{4L_i}\left(4 - k^2 L_i^2\right) \\ q_i &= \frac{1}{4L_i}\left(-4 - k^2 L_i^2\right) \end{aligned}\right\} \quad i = 1 \ldots n. \tag{9}$$

where L_i are the parameters that determine the accuracy of the approximation.

The error in the approximation (7) is typically analyzed through the so-called reflection coefficient, which has been shown to be (for details, see [4])

$$R = \prod_{i=1}^{n} \left| \frac{\mathscr{K}_1 - p_i}{\mathscr{K}_1 + p_i} \right|^2 . \tag{10}$$

If $R = 0$, then the approximation is exact, and the deviation from zero indicates magnitude of error in the approximation; smaller the value of R, better the approximation. So from (10) and (9), it is clear that the accuracy of proposed approximation hinges on the choice of L_i.

In general, L_i are chosen to be complex or imaginary to better approximate the DtN map for propagating wave modes and are chosen to be real when evanescent modes are important. While the parameters L_i can be optimized using the concepts discussed in [5], in this paper we choose L_i based on the OSM parameters (see Sect. 4).

Implementation of PMDL. While the rational form of the PMDL approximation in (8) is useful for analysis, the following matrix form proves to be useful for implementation:

$$\begin{bmatrix} \mathscr{S}_n^{pmdl} u_b \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} p_1 & q_1 & 0 & \cdots & 0 \\ q_1 & p_1 + p_2 & q_2 & & 0 \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & & q_{n-1} & p_{n-1} + p_n & q_n \\ 0 & \cdots & 0 & q_n & p_n \end{bmatrix} \begin{bmatrix} u_b \\ u_{a,1} \\ u_{a,2} \\ \vdots \\ u_{a,n-1} \end{bmatrix}, \tag{11}$$

where p_i, q_i are given by (9) and $u_{a,i}$ are auxiliary variables that are introduced to facilitate the implementation and have no direct physical relevance to the problem. The equivalence between (8) and (11) can be easily seen by eliminating the auxiliary dof $u_{a,i}$ from (11) to recover (8). The matrix form of PMDL enables an easy implementation of the rational approximation as a simple tri-diagonal matrix.

PMDL, a link between Rational ABCs and Perfectly Matched Layers. While the matrix form of the PMDL approximation in (11) is based on the rational approximation in (8), it is intimately linked to impedance-preserving discretization of PML proposed in [4]. Unlike PML, the impedance is preserved even after discretization and thus the approximation is named perfectly matched discrete layers, PMDL. This link is substantial in that it provides a way to derive and easily implement PMDL approximations for more complex cases such as corners [4] and anisotropic elasticity [6].

The ease of implementation of PMDL is in fact the impetus behind proposed method. As implied by (10), the accuracy of approximation can be easily increased by adding auxiliary variables, which is equivalent to adding lines of nodes parallel to the interface. As will be shown in Sect. 4, addition of just one auxiliary variable, which has minimal increase in computational cost per iteration, significantly reduces the convergence factor and the number of iterations needed.

Implementation of the PMDL-Schwarz method. The proposed PMDL-Schwarz method is essentially the Schwarz Alternating method with the operator Λ_1 chosen to be the DtN map obtained using PMDL, i.e., $\Lambda_1 = \mathscr{S}_n^{pmdl}$ where $\mathscr{S}_n^{pmdl}$ is given by (11). Thus the interface condition in (3) for Ω_1 can be written as

$$\frac{\partial}{\partial \mathbf{n_1}}(u_1^{j+1} - u_2^j) + \mathscr{S}_n^{pmdl}(u_1^{j+1} - u_2^j) = 0\,. \tag{12}$$

Substituting (11) in (12), we get the PMDL-Schwarz formulation as

$$\begin{bmatrix} \frac{\partial u_1^{j+1}}{\partial \mathbf{n_1}} \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \begin{bmatrix} p_1 & q_1 & 0 & \cdots & 0 \\ q_1 & p_1+p_2 & q_2 & & 0 \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & & q_{n-1} & p_{n-1}+p_n & q_n \\ 0 & \cdots & 0 & q_n & p_n \end{bmatrix} \begin{bmatrix} u_1^{j+1} \\ u_{a,1} \\ u_{a,2} \\ \vdots \\ u_{a,n-1} \end{bmatrix} = \begin{bmatrix} -\frac{\partial u_2^j}{\partial \mathbf{n_2}} + p_1 u_2^j \\ q_1 u_2^j \\ 0 \\ \vdots \\ 0 \end{bmatrix}. \tag{13}$$

Note that the formulation of the interface condition for Ω_2 can be derived in an identical manner and hence is not repeated here.

4 Comparison Between OSM and PMDL-Schwarz Methods

In this section, we compare the performance of OSM and PMDL-Schwarz method both theoretically (using convergence factors) and in a numerical simulation involving multiple domains and closed boundaries.

Convergence Factors: Consider the stiffness approximation of the second-order OSM (see [3]),

$$\mathscr{S}_{osm} = \frac{ab - \omega^2}{a+b} + \frac{1}{a+b}k_y^2\,. \tag{14}$$

Substituting $\Lambda_1 = \Lambda_2 = \mathscr{S}_{osm}$ in (5), we get the convergence factor of OSM to be

$$\rho_{osm} = \left| \frac{ab + k_y^2 - \omega^2 + \mathrm{i}\,(a+b)\sqrt{\omega^2 - k_y^2}}{ab + k_y^2 - \omega^2 - \mathrm{i}\,(a+b)\sqrt{\omega^2 - k_y^2}} \right|^2 .$$

To compare, we use a two-layer PMDL-Schwarz method with $L_1 = 2/a$, and $L_2 = 2/b$, where a, b are the OSM parameters in (14). The stiffness approximation of the two-layer PMDL-Schwarz method is then given by

$$\mathscr{S}_n^{pmdl} = p_2 - \frac{q_2^2}{p_2 + p_1},$$
$$p_2 = \frac{1}{L_2} - \frac{(\omega^2 - k_y^2) L_2}{4}, \quad q_2 = -\frac{1}{L_2} - \frac{(\omega^2 - k_y^2) L_2}{4},$$
$$p_1 = \frac{1}{L_1} - \frac{(\omega^2 - k_y^2) L_1}{4}.$$

Substituting $\Lambda_1 = \Lambda_2 = \mathscr{S}_n^{pmdl}$ in (5), we get the convergence factor of PMDL-Schwarz that can be simplified to

$$\rho_{pmdl} = \left(\left| \frac{ab + k_y^2 - \omega^2 + \mathrm{i}(a+b)\sqrt{\omega^2 - k_y^2}}{ab + k_y^2 - \omega^2 - \mathrm{i}(a+b)\sqrt{\omega^2 - k_y^2}} \right|^2 \right)^2.$$

Clearly $\rho_{pmdl} = \rho_{osm}^2$, and so the parameters of PMDL-Schwarz are chosen such that its convergence factor is the square of that of OSM and the method performs uniformly better over the entire range of wavenumbers k_y.

It is easy to numerically verify the above result for the model problem (1a), with the domain Ω decomposed into two semi-infinite layers. We take $a = 20.741\,\mathrm{i}$ and $b = 47.071$ to be the OSM parameters as these were shown in [3] to be optimal over the allowed wavenumber range $k_y \in [\pi, 60\pi]$. Figure 1a compares the convergence factors of OSM and PMDL-Schwarz method (with $L_1 = 2/a$ and $L_2 = 2/b$) and shows clearly that the proposed method performs better over the entire range of wavenumbers for a slightly increased computational cost (there is only one auxiliary variable introduced, which is similar to one line of nodes in 2-D).

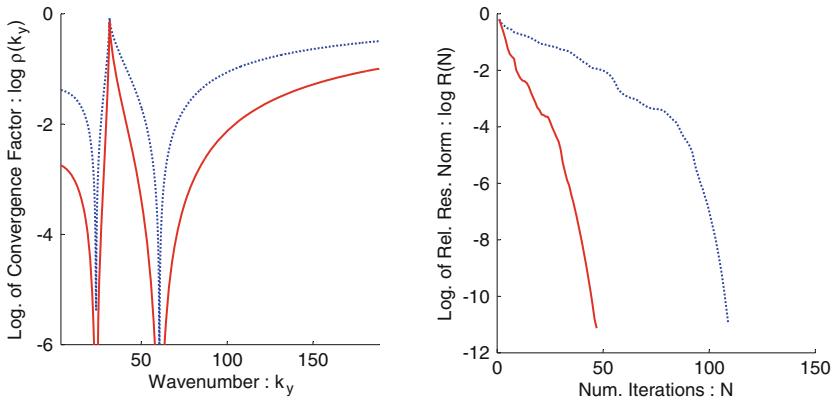


Fig. 1. Comparison between OSM (*dotted line*) and PMDL-Schwarz method (*solid line*). **(a)** Convergence Factor. **(b)** Convergence of Solution

Numerical Example: In this example, Eq. (1a) is solved on a square domain ($\Omega \equiv [0,1] \times [0,1]$) with $\omega = 10\pi$ and a point source $f = 1/2$ is applied at $(0, 0.5)$. Homogeneous Neumann boundary condition is applied on the left ($x = 0$), Dirichlet condition at the top ($y = 1$) and bottom ($y = 0$), and an ABC on the right ($x = 1$). The computational domain is discretized using 60 bilinear finite elements along each direction. The domain is decomposed into nine subdomains with three subdomains along each dimension. The convergence plot is shown in Fig. 1b. As expected, the PMDL-Schwarz method converges twice as fast as the conventional OSM.

5 Discussion

We proposed a Schwarz method for Helmholtz equation based on the concepts of perfectly matched discrete layers (PMDL), a recently developed absorbing boundary condition that is related to the higher order rational approximations and the Perfectly Matched Layers. By examining the convergence factor and with the help of a numerical example, PMDL-Schwarz method is shown to converge faster than existing Optimized Schwarz Methods. Although not treated in this paper, it is important to mention that the PMDL is not just limited to the Helmholtz equation, but also to more complicated vector equations such as the elastic and electromagnetic wave equations. Thus, it is expected that the PMDL-Schwarz method would provide accelerated convergence in frequency domain computations in these contexts. Furthermore, as Waveform Relaxation Method in time domain share similar ideas with OSM (see e.g. [2]), PMDL ideas can also be used to improve the convergence of existing waveform relaxation methods. These extensions are subjects of ongoing research.

Bibliography

[1] JP Berenger. A perfectly matched Layer for the absorption of electromagnetic-waves. *Journal of Computational Physics*, 114(2):185–200, OCT 1994. ISSN 0021-9991. doi: {10.1006/jcph.1994.1159}.

[2] M. J. Gander, L. Halpern, and F. Nataf. Optimal convergence for overlapping and non-overlapping Schwarz waveform relaxation. In *Eleventh International Conference on Domain Decomposition Methods (London, 1998)*, pages 27–36 (electronic). DDM.org, Augsburg, 1999.

[3] Martin J. Gander, Frédéric Magoulès, and Frédéric Nataf. Optimized Schwarz methods without overlap for the Helmholtz equation. *SIAM J. Sci. Comput.*, 24(1):38–60 (electronic), 2002. ISSN 1064-8275. doi: 10.1137/S1064827501387012.

[4] Murthy N. Guddati and Keng-Wit Lim. Continued fraction absorbing boundary conditions for convex polygonal domains. *Internat. J. Numer. Methods Engrg.*, 66(6):949–977, 2006. ISSN 0029-5981. doi: 10.1002/nme.1574.

[5] David Ingerman, Vladimir Druskin, and Leonid Knizhnerman. Optimal finite difference grids and rational approximations of the square root. I. Elliptic problems. *Comm. Pure Appl. Math.*, 53(8):1039–1066, 2000. ISSN 0010-3640. doi: 10.1002/1097-0312(200008)53:8<1039::AID-CPA4>3.3.CO;2-9.

[6] Siddharth Savadatti and Murthy N. Guddati. Absorbing boundary conditions for scalar waves in anisotropic media. part 1: Time harmonic modeling. *Journal of Computational Physics*, 229(19):6696 – 6714, 2010. ISSN 0021-9991. doi: 10.1016/j.jcp.2010.05.018.

[7] Andrea Toselli and Olof Widlund. *Domain decomposition methods—algorithms and theory*, volume 34 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin, 2005. ISBN 3-540-20696-5.

[8] SV Tsynkov. Numerical solution of problems on unbounded domains. A review. *Applied Numerical Mathematics*, 27(4):465–532, 1998. ISSN 0168-9274. doi: {10.1016/S0168-9274(98)00025-7}.

A Domain Decomposition Solver for the Discontinuous Enrichment Method for the Helmholtz Equation

Charbel Farhat[1], Radek Tezaur[1], and Jari Toivanen[1]

Department of Aeronautics & Astronautics, Stanford University, Mail Code 3035, Stanford, CA 94305, U.S.A. cfarhat@stanford.edu, rtezaur@stanford.edu, toivanen@stanford.edu

1 Introduction

The discontinuous enrichment method (DEM) [4] for the Helmholtz equation approximates the solution as a sum of a piecewise polynomial continuous function and element-wise supported plane waves [5]. A weak continuity of the plane wave part is enforced using Lagrange multipliers. The plane wave enrichment improves the accuracy of solutions considerably. In the mid-frequency range, severalfold savings in terms of degrees of freedom over comparable higher order polynomial discretizations have been observed, which translates into even larger savings in compute time [6, 9]. The partition of unity method [8] and the ultra weak variational formulation [1] also employ plane waves in the construction of discretizations. It was shown recently in [10] that DEM without the polynomial field is computationally more efficient than these methods.

So far only direct solution methods have been used with DEM. This paper describes an iterative domain decomposition method which will enable to solve much larger problems with DEM. The method is a generalization of the FETI-H version [3] of the FETI method [2] and the domain decomposition method for DEM without the polynomial part described in [7]. It is based on a non-overlapping decomposition of the domain into subdomains. On the subdomain interfaces Lagrange multipliers are introduced to enforce the continuity of the polynomial part strongly and the continuity of the enrichment weakly. An efficient iterative solution procedure with a two-level preconditioner resembling that of the FETI-H method is constructed for the Lagrange multipliers on the interfaces between the subdomains.

2 Problem Formulation and Discretization

The solution $u \in H^1(\Omega)$ of a Helmholtz problem modeling acoustic scattering from a rigid obstacle, for example, satisfies the equations

R. Bank et al. (eds.), *Domain Decomposition Methods in Science and Engineering XX*, Lecture Notes in Computational Science and Engineering 91,
DOI 10.1007/978-3-642-35275-1_23,

$$\begin{aligned} -\Delta u - k^2 u &= f && \text{in } \Omega \\ \frac{\partial u}{\partial \nu} &= g_1 && \text{on } \Sigma_1 \\ \frac{\partial u}{\partial \nu} &= iku + g_2 && \text{on } \Sigma_2, \end{aligned} \tag{1}$$

where k is the wavenumber, Σ_1 is the boundary of a sound-hard scatterer, Σ_2 is the far-field boundary, and ν denotes the unit outward normal.

Let the domain Ω be split into n_e elements, $\Omega = \cup_{e=1}^{n_e} \Omega_e$. In DEM, the solution is sought in the form $u = u^P + u^E$, where u^P is a standard continuous piecewise polynomial finite element function, and u^E is an enrichment function discontinuous across element interfaces. A weak inter-element continuity of the solution is enforced by Lagrange multipliers λ^E. The following hybrid variational formulation is used: Find $u \in \mathcal{V}$ and $\lambda^E \in \mathcal{W}^E$ such that

$$\begin{aligned} a(u,v) + b(\lambda^E, v) &= r(v) && \forall v \in \mathcal{V} \\ b(\mu^E, u) &= 0 && \forall \mu^E \in \mathcal{W}^E. \end{aligned}$$

The forms a, b, and r are defined by

$$\begin{aligned} a(u,v) &= \int_\Omega (\nabla u \cdot \nabla v - k^2 uv) d\Omega - \int_{\Sigma_2} ikuv\, d\Gamma, \\ b(\lambda^E, v) &= \sum_{e=1}^{n_e} \sum_{e'=1}^{e-1} \int_{\Gamma_{e,e'}} \lambda^E \left(v_{\Omega_e'} - v|_{\Omega_e} \right) d\Gamma, \quad \text{and} \\ r(v) &= \int_\Omega f v\, d\Omega + \int_{\Sigma_1} g_1 v\, d\Gamma + \int_{\Sigma_2} g_2 v\, d\Gamma, \end{aligned}$$

where $\Gamma_{e,e'} = \partial\Omega_e \cap \partial\Omega_{e'}$. For the considered discretization, the space $\mathcal{V}$ consists of functions of the form $u = u^P + u^E$, where u^E is a superposition of n_θ planar waves, i.e.

$$u^E(\mathbf{x}) = \sum_{p=1}^{n_\theta} e^{ik\theta_p \cdot \mathbf{x}} u^E_{e,p}, \qquad \mathbf{x} \in \Omega_e.$$

In two dimensions, $\theta_p = (\cos\vartheta_p, \sin\vartheta_p)^T$, $\vartheta_p = 2\pi(p-1)/n_\theta, p = 1, \ldots, n_\theta$. The Lagrange multipliers space $\mathcal{W}^E$ is then chosen using functions of the form

$$\lambda^E(\mathbf{x}) = \sum_{p=1}^{n_\lambda} e^{ik\eta_p \tau_{e,e'} \cdot \mathbf{x}} \lambda_{e,e',p}, \qquad \mathbf{x} \in \Gamma_{e,e'},$$

where $\tau_{e,e'}$ is a unit tangent vector and η_p is a scalar. This choice yields a family of quadrilateral elements, denoted by Q-n_θ-n_λ. In particular, the elements Q-8-2 and Q-16-4 used in the numerical experiments in this paper use $\eta_1 = -\eta_2 = 0.5$ and $\{\eta_p\}_{p=1}^4 = \{\pm 0.2, \pm 0.75\}$, respectively. For details on stability, implementation, and accuracy, the reader is referred to [5, 6].

3 Domain Decomposition Formulation

The elements are divided into n^d disjoint subsets E^j defining subdomains Ω^j such that $\bar{\Omega}^j = \cup_{e \in E^j} \bar{\Omega}_e$. Subdomain problems are given by regularized bilinear forms

$$\begin{aligned}\widetilde{a}^j(u^j, v^j) = &\int_{\Omega^j} (\nabla u^j \cdot \nabla v^j - k^2 u^j v^j) d\Omega - \int_{\Sigma_2 \cap \partial\Omega^j} iku^j v^j \, d\Gamma \\ &- \gamma \sum_{\substack{j'=1 \\ j' \neq j}}^{n^d} \int_{\Gamma^{j,j'}} s^{j,j'} iku^j v^j \, d\Gamma,\end{aligned}$$

where $\Gamma^{j,j'} = \partial\Omega^j \cap \partial\Omega^{j'}$. The functions u^j and v^j belong to the restriction of $\mathcal{V}$ into Ω^j and the last term ensures the subdomain problems cannot be singular; for details see [7]. The coefficients $s^{j,j'}$ are chosen so that the regularization terms cancel out for a continuous function. The continuity of the polynomial part of the solution $\tilde{u}^P = \sum_{j=1}^{n^d} u^{P,j}$ across the subdomain interfaces is enforced using a Lagrange multiplier λ^P. For this purpose, a bilinear form

$$c(\lambda^P, \tilde{v}) = \sum_{j=1}^{n^d} \sum_{j'=1}^{j-1} \sum_{l} \lambda^P_{j,j',l} \left(\tilde{v}^P|_{\Omega^{j'}} - \tilde{v}^P|_{\Omega^j} \right) (\mathbf{x}_{j,j',l})$$

is defined, where $\mathbf{x}_{j,j',l}$ is the location of the lth mesh node on $\Gamma^{j,j'}$. The mesh nodes are given by the Lagrange interpolation points of the piecewise polynomial functions. The domain decomposition formulation then reads:

Find $\tilde{u} \in \widetilde{\mathcal{V}}$, λ^E, and λ^P such that

$$\begin{aligned} \widetilde{a}(\tilde{u}, \tilde{v}) + b(\lambda^E, \tilde{v}) + c(\lambda^P, \tilde{v}) &= \widetilde{r}(\tilde{v}) && \forall \tilde{v} \in \widetilde{\mathcal{V}} \\ b(\mu^E, \tilde{u}) &= 0 && \forall \mu^E \in \mathcal{W}^E \\ c(\mu^P, \tilde{u}) &= 0 && \forall \mu^P \in \mathcal{W}^P, \end{aligned} \tag{2}$$

where $\widetilde{\mathcal{V}}$ is spanned by $\sum_{j=1}^{n^d} v_j$, $\widetilde{a}(\tilde{u}, \tilde{v}) = \sum_{j=1}^{n^d} a^j(u^j, v^j)$, and $\widetilde{r}$ is the sum of subdomain contributions of r.

4 Linear Systems and Condensations

The formulation (2) leads to the saddle point system of linear equations

$$\begin{pmatrix} \mathbf{rA}^{PP} & \mathbf{rA}^{PE} & 0 & \mathbf{C}^{PL} \\ \mathbf{rA}^{EP} & \mathbf{rA}^{EE} & \mathbf{B}^{EL} & 0 \\ 0 & \mathbf{B}^{LE} & 0 & 0 \\ \mathbf{C}^{LP} & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \boldsymbol{u}^P \\ \boldsymbol{u}^E \\ \boldsymbol{\lambda}^E \\ \boldsymbol{\lambda}^P \end{pmatrix} = \begin{pmatrix} \mathbf{r}^P \\ \mathbf{r}^E \\ 0 \\ 0 \end{pmatrix}, \tag{3}$$

where the superscripts P, E, and L refer to the polynomial part, the enrichment part, and the Lagrange multiplier, respectively, and $\boldsymbol{u}^P, \boldsymbol{u}^E, \boldsymbol{\lambda}^E, \boldsymbol{\lambda}^P$ are vectors of the subdomain-by-subdomain polynomial degrees of freedom (depicted by black dots in Fig. 1), the element-by-element enrichment degrees of freedom (magenta arrows), the enrichment element-to-element continuity Lagrange multipliers (red arrows), and the polynomial subdomain-to-subdomain continuity Lagrange multipliers (black arrows), respectively. The enrichment unknowns $\boldsymbol{u}^E$ can be condensed out on the element level (Fig. 1 top and left) to obtain

$$\begin{pmatrix} \bar{\mathbf{rA}} & \bar{\mathbf{B}}^T & \bar{\mathbf{C}}^T \\ \bar{\mathbf{B}} & \bar{\mathbf{D}} & 0 \\ \bar{\mathbf{C}} & 0 & 0 \end{pmatrix} \begin{pmatrix} \boldsymbol{u}^P \\ \boldsymbol{\lambda}^E \\ \boldsymbol{\lambda}^P \end{pmatrix} = \begin{pmatrix} \bar{\mathbf{r}} \\ \bar{\boldsymbol{\mu}} \\ 0 \end{pmatrix}, \tag{4}$$

where

$$\begin{aligned} \bar{\mathbf{rA}} &= \mathbf{rA}^{PP} - \mathbf{rA}^{PE} \left(\mathbf{rA}^{EE}\right)^{-1} \mathbf{rA}^{EP}, & \bar{\mathbf{B}} &= -\mathbf{B}^{LE} \left(\mathbf{rA}^{EE}\right)^{-1} \mathbf{rA}^{EP}, \\ \bar{\mathbf{C}} &= \mathbf{C}^{LP}, & \bar{\mathbf{D}} &= -\mathbf{B}^{LE} \left(\mathbf{rA}^{EE}\right)^{-1} \mathbf{B}^{EL}, \\ \bar{\mathbf{r}} &= \mathbf{r}^P - \mathbf{rA}^{PE} \left(\mathbf{rA}^{EE}\right)^{-1} \mathbf{r}^E, & \bar{\boldsymbol{\mu}} &= -\mathbf{B}^{LE} \left(\mathbf{rA}^{EE}\right)^{-1} \mathbf{r}^E. \end{aligned}$$

The enrichment Lagrange multipliers $\boldsymbol{\lambda}^E$ can be divided into two parts—those on the boundaries between the subdomains and those inside the subdomains, denoted by the subscript B and I, respectively. The system (4) can then be written in the block form

$$\begin{pmatrix} \bar{\mathbf{rA}} & \bar{\mathbf{B}}_{II}^T & \bar{\mathbf{B}}_{BB}^T & \bar{\mathbf{C}}^T \\ \bar{\mathbf{B}}_{II} & \bar{\mathbf{D}}_{II} & \bar{\mathbf{D}}_{IB} & 0 \\ \bar{\mathbf{B}}_{BB} & \bar{\mathbf{D}}_{BI} & \bar{\mathbf{D}}_{BB} & 0 \\ \bar{\mathbf{C}} & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \boldsymbol{u}^P \\ \boldsymbol{\lambda}_I^E \\ \boldsymbol{\lambda}_B^E \\ \boldsymbol{\lambda}^P \end{pmatrix} = \begin{pmatrix} \bar{\mathbf{r}} \\ \bar{\boldsymbol{\mu}}_I \\ \bar{\boldsymbol{\mu}}_B \\ 0 \end{pmatrix}.$$

Finally, the elimination on the subdomain level of the unknowns $\boldsymbol{u}^P$ and the interior

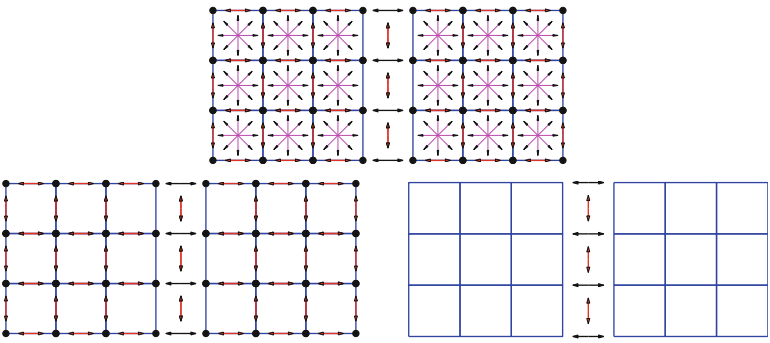


Fig. 1. 2×1 domain decomposition of a DEM discretization with bilinear polynomials and Q-8-2 elements resulting in the system (3) (*top*); variables left after condensation of enrichment dofs (4) (*left*); and elimination of the subdomain interior dofs (5) (*right*)

enrichment Lagrange multipliers $\boldsymbol{\lambda}_I^E$ gives the Schur complement system (cf. Fig. 1 right)

$$\mathbf{F}\begin{pmatrix}\boldsymbol{\lambda}_B^E\\ \boldsymbol{\lambda}^P\end{pmatrix} = \mathbf{b}. \tag{5}$$

It is noted that the matrix $\mathbf{F}$ is a sum of subdomain matrices. Once the Lagrange multipliers $\boldsymbol{\lambda}_B^E$ and $\boldsymbol{\lambda}^P$ have been solved from (5), the rest of the unknowns is recovered by post-processing, first to obtain $\boldsymbol{u}^P$ and $\boldsymbol{\lambda}_I^E$, then to obtain $\boldsymbol{u}^E$.

5 Preconditioning

The system (5) is solved efficiently using a Krylov iterative method with a two-level preconditioner which is a generalization of those described in [3, 7].

Here, the subdomain preconditioners are based on the bilinear forms

$$\hat{a}^j(u^j, v^j) = \int_{\Omega^j} (\nabla u^j \cdot \nabla v^j - k^2 u^j v^j) d\Omega - \int_{\partial\Omega^j \setminus \Sigma_1} iku^j v^j \, d\Gamma,$$

$$\hat{b}^j(\lambda^E, v^j) = \sum_{e\in E^j} \sum_{e'=e+1}^{n_e} \int_{\Gamma_{e,e'}} \lambda^E v|_{\Omega_e} d\Gamma - \sum_{e\in E^j} \sum_{e'=1}^{e-1} \int_{\Gamma_{e,e'}} \lambda^E v|_{\Omega_e} d\Gamma, \quad \text{and}$$

$$\hat{c}^j(\lambda^P, v^j) = \sum_{j'=j+1}^{n^d} \sum_{l} \lambda^P_{j,j',l} v^P|_{\Omega^j}(\mathbf{x}_{j,j',l}) - \sum_{j'=1}^{j-1} \sum_{l} \lambda^P_{j,j',l} v^P|_{\Omega^j}(\mathbf{x}_{j,j',l}).$$

Repeating the same steps described above for obtaining $\mathbf{F}$ in (5) but with matrices based on $\hat{a}^j$, and restricting the resulting matrix to the unknowns corresponding to the interfaces of the subdomain Ω^j, a matrix denoted by $\mathbf{F}^j$ is obtained (cf. [7]). An additive subdomain-by-subdomain preconditioner is then defined by

$$\mathbf{K} = \sum_{j=1}^{n^d} \left(\mathbf{R}^j\right)^T \left(\mathbf{F}^j\right)^{-1} \mathbf{R}^j,$$

where $\mathbf{R}^j$ is the restriction on the interfaces associated with Ω^j. Linear systems with $\mathbf{F}^j$ can be solved efficiently using an $\mathbf{LU}$ decomposition.

The system (5) is solved iteratively on the orthogonal complement of a coarse space spanned by the columns of a matrix $\mathbf{Q}$ (cf. [3, 7]). A projector to the orthogonal complement of the coarse space is given by

$$\mathbf{P} = \mathbf{I} - \mathbf{Q}(\mathbf{Q}^T\mathbf{F}\mathbf{Q})^{-1}\mathbf{Q}^T\mathbf{F}.$$

The solution $\boldsymbol{\lambda} = [\boldsymbol{\lambda}_B^E, \boldsymbol{\lambda}^P]^T$ of (5) can be decomposed into two parts $\boldsymbol{\lambda} = \boldsymbol{\lambda}^0 + \mathbf{P}\boldsymbol{\lambda}^1$, where $\boldsymbol{\lambda}^0 = \mathbf{Q}(\mathbf{Q}^T\mathbf{F}\mathbf{Q})^{-1}\mathbf{Q}^T\mathbf{b}$ and $\boldsymbol{\lambda}^1$ satisfies

$$\mathbf{P}^T\mathbf{F}\boldsymbol{\lambda}^1 = \mathbf{P}^T\mathbf{b}.$$

Including the preconditioner $\mathbf{K}$ leads to the following equation

$$\mathbf{PKP}^T\mathbf{F}\boldsymbol{\lambda}^1 = \mathbf{PKF}\boldsymbol{\lambda}^1 = \mathbf{PKP}^T\mathbf{b},$$

which is solved by GMRES.

The coarse space is based on plane waves propagating in n_q uniformly distributed directions. Each set of n_q plane waves are supported by one subdomain interface $\Gamma^{j,j'}$ and their normal derivatives on the interface are approximated using an L^2-projection into the space of Lagrange multipliers giving rise to n_q columns of $\mathbf{Q}$. Currently, the coarse space acts only on the interface enrichment Lagrange multipliers λ_B^E. The maximum dimension of the coarse space is $n_q n_i$, where n_i is the number of nonzero measure interfaces $\Gamma^{j,j'}$. A **QR** factorization is used to remove nearly linearly dependent vectors. More details are given in Sect. 3.4 of [7].

6 Numerical Results

The model problem considered here is given by (1) with the computational domain $\Omega = \{\mathbf{x} \in \mathbb{R}^2 : 1 < \|x\| < 2\}$, and the boundaries $\Gamma_1 = \{\mathbf{x} \in \mathbb{R}^2 : \|x\| = 1\}$ and $\Gamma_2 = \{\mathbf{x} \in \mathbb{R}^2 : \|x\| = 2\}$. The right-hand side function and the boundary functions are chosen as

$$\begin{aligned} f(\mathbf{x}) &= (-\Delta - k^2)(x_1^2 + x_2^2) = -4 - k^2(x_1^2 - x_2^2), \\ g_1(\mathbf{x}) &= -\frac{\partial e^{-ikx_1}}{\partial \nu} + \frac{\partial (x_1^2 + x_2^2)}{\partial \nu} = -ikx_1 e^{ikx_1} - 2(x_1^2 + x_2^2), \quad \text{and} \\ g_2(\mathbf{x}) &= \frac{\partial (x_1^2 + x_2^2)}{\partial \nu} - ik(x_1^2 + x_2^2) = (1 - ik)(x_1^2 + x_2^2). \end{aligned}$$

The solution is a sum of that given by the scattering of the plane wave e^{-ikx_1} by a sound-hard disk inside Γ_1 and the polynomial $x_1^2 + x_2^2$. Two wavenumbers, $k = 8\pi$ and 16π are considered, in which case the diameter of the scatterer is 8 and 16 wavelengths, respectively. The solution at $k = 16\pi$ is shown in Fig. 2. Meshes of 96×8 ($k = 8\pi$) and 192×16 ($k = 16\pi$) elements result in two elements per wavelength in the radial direction.

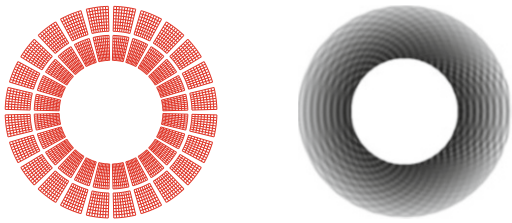


Fig. 2. The 24×2 domain decomposition for the 192×16 mesh (*left*) and the real part of the solution at $k = 16\pi$ (*right*)

Table 1. Results for the 96×8 mesh with the wavenumber $k = 8\pi$.

		12 x 1 subdomains			24 x 2 subdomains			
			$n_q = 0$	$n_q = 8$		$n_q = 0$	$n_q = 8$	
poly	enrich	N	iter.	iter.	N	iter.	iter.	error
Q_1	none	108	49		336	213		0.683405
Q_2	none	204	33		624	195		0.141341
none	Q-8-2	192	35	31	576	163	7	0.438341
Q_1	Q-8-2	300	34	31	912	184	28	0.004677
Q_2	Q-8-2	396	34	31	1200	206	48	0.004472
none	Q-16-4	384	35	30	1152	151	39	0.019767
Q_1	Q-16-4	492	36	31	1488	160	54	0.000024
Q_2	Q-16-4	588	36	31	1776	176	73	0.000013

Table 2. Results for the 192×16 mesh with the wavenumber $k = 16\pi$.

		12 x 1 subdomains			24 x 2 subdomains			
			$n_q = 0$	$n_q = 16$		$n_q = 0$	$n_q = 16$	
poly	enrich	N	iter.	iter.	N	iter.	iter.	error
Q_1	none	204	79		624	350		0.568750
Q_2	none	396	40		1200	368		0.174451
none	Q-8-2	384	44	34	1152	264	16	0.478914
Q_1	Q-8-2	588	42	34	1776	281	31	0.007441
Q_2	Q-8-2	780	42	34	2352	295	56	0.007826
none	Q-16-4	768	42	33	2304	233	42	0.021694
Q_1	Q-16-4	972	42	35	2928	238	52	0.000011
Q_2	Q-16-4	1164	42	33	3504	253	123	0.000010

Bilinear (Q_1) and biquadratic (Q_2) bases are used for the polynomial part $\boldsymbol{u}^P$. Q-8-2 and Q-16-4 elements are used for the enrichment $\boldsymbol{u}^E$ and its Lagrange multipliers $\boldsymbol{\lambda}^E$. The domain is decomposed into 12×1 and 24×2 subdomains (Fig. 2). The GMRES iterations are terminated once the norm of the residual is reduced by 10^{-8}. Tables 1 and 2 summarize the performance results obtained for various element types. In these tables, N is the size of the system (5), i.e. the number of Lagrange multipliers enforcing continuity between subdomains. The error is the relative l_2 error of the averaged nodal values with respect to the analytical solution of the problem.

The errors in the last column of Tables 1 and 2 clearly show the benefit of discretizations with both polynomial and enrichment fields for this problem. The combined discretizations increase the accuracy by at least two orders of magnitude. The iteration counts without a coarse space ($n_q = 0$) are roughly the same for all discretizations and not quite satisfactory for the 24×2 decomposition. However, these are reduced substantially when the coarse space is added.

Bibliography

[1] Olivier Cessenat and Bruno Despres. Application of an ultra weak variational formulation of elliptic PDEs to the two-dimensional Helmholtz problem. *SIAM J. Numer. Anal.*, 35(1):255–299, 1998.

[2] Charbel Farhat and Francoise-Xavier Roux. A method of finite element tearing and interconnecting and its parallel solution algorithm. *Internat. J. Numer. Meths. Engrg.*, 32(6):1205–1227, 1991.

[3] Charbel Farhat, Antonini Macedo, and Michel Lesoinne. A two-level domain decomposition method for the iterative solution of high frequency exterior Helmholtz problems. *Numer. Math.*, 85(2):283–308, 2000.

[4] Charbel Farhat, Isaac Harari, and Leopoldo P. Franca. The discontinuous enrichment method. *Comput. Methods Appl. Mech. Engrg.*, 190(48):6455–6479, 2001.

[5] Charbel Farhat, Isaac Harari, and Ulrich Hetmaniuk. A discontinuous Galerkin method with Lagrange multipliers for the solution of Helmholtz problems in the mid-frequency regime. *Comput. Methods Appl. Mech. Engrg.*, 192(11–12): 1389–1419, 2003.

[6] Charbel Farhat, Radek Tezaur, and Paul Weidemann-Goiran. Higher-order extensions of a discontinuous Galerkin method for mid-frequency Helmholtz problems. *Internat. J. Numer. Methods Engrg.*, 61(11):1938–1956, 2004.

[7] Charbel Farhat, Radek Tezaur, and Jari Toivanen. A domain decomposition method for discontinuous Galerkin discretizations of Helmholtz problems with plane waves and Lagrange multipliers. *Internat. J. Numer. Methods Engrg.*, 78 (13):1513–1531, 2009.

[8] Jens M. Melenk and Ivo Babuška. The partition of unity finite element method: basic theory and applications. *Comput. Methods Appl. Mech. Engrg.*, 139(1–4): 289–314, 1996.

[9] Radek Tezaur and Charbel Farhat. Three-dimensional discontinuous Galerkin elements with plane waves and Lagrange multipliers for the solution of mid-frequency Helmholtz problems. *Internat. J. Numer. Methods Engrg.*, 66(5): 796–815, 2006.

[10] Dalei Wang, Radek Tezaur, Jari Toivanen, and Charbel Farhat. Overview of the discontinuous enrichment method, the ultra-weak variational formulation, and the partition of unity method for acoustic scattering in the medium frequency regime and performance comparisons. *Internat. J. Numer. Methods Engrg.*, 89 (4):403–417, 2012.

Domain Decomposition Methods for the Helmholtz Equation: A Numerical Investigation

Martin J. Gander and Hui Zhang

University of Geneva martin.gander@unige.ch, hui.zhang@unige.ch

1 Introduction

We are interested in solving the Helmholtz equation

$$\begin{cases} -\triangle u(x,y,z) - k^2(x,y,z)\, u(x,y,z) = g(x,y,z), & (x,y,z) \in \Omega, \\ \partial_n u(x,y,z) - \mathbf{i}k(x,y,z)\, u(x,y,z) = 0, & (x,y,z) \in \partial\Omega, \end{cases} \tag{1}$$

where $k := 2\pi f/c$ is the wavenumber with frequency $f \in \mathbf{R}$ and $c := c(x,y,z)$ is the velocity of the medium, which varies in space. The geophysical model SEG–SALT is used as a benchmark problem on which we will test some existing domain decomposition methods in this paper. In this model, the domain Ω is defined as $(0,13{,}520) \times (0,13{,}520) \times (0,4{,}200)\,\mathrm{m}^3$, the velocity is described as piecewise constants on $676 \times 676 \times 210$ cells and varies from 1,500 to 4,500 m/s, and the source g is a Dirac function at the point $(6{,}000, 6{,}760, 10)$.

To discretize the problem (1) on a coarser mesh, the velocity is sub-sampled to less number of cells such that every cell has a constant velocity and contains one or more mesh elements. Then the problem (1) is discretized with $Q1$ finite elements (i.e. trilinear local basis functions on brick elements).

We first test the direct solver $A\backslash b$ in Matlab; the results are listed in Table 1 where nw is the number of wavelength along the x-direction at the lowest velocity. At $f = 2$, the direct solver runs out of memory after 6 h on a computer with 64 GB of memory. The inefficiency in both memory and time of the direct solver for large scale problems calls for cheaper iterative methods. For a review of current iterative methods for the Helmholtz equation, we refer to [6]. In this work, we focus on domain decomposition methods which are easily parallelized.

2 Overview of Some Existing Methods

Due to the indefiniteness of the Helmholtz equation, the classical Schwarz method with Dirichlet transmission conditions fails to converge. As a remedy, [5] introduced

R. Bank et al. (eds.), *Domain Decomposition Methods in Science and Engineering XX*, Lecture Notes in Computational Science and Engineering 91,
DOI 10.1007/978-3-642-35275-1_24,

Table 1. Test of the direct solver (backslash in Matlab)

f	1/4	1/2	1	2
nw	2.25	4.5	9	18
mesh	$24\times24\times8$	$48\times48\times16$	$96\times96\times32$	$192\times192\times64$
CPU	1.28s	27.51s	829.91s	$>$ 6h

first-order absorbing transmission conditions to replace the Dirichlet transmission conditions. This type of interface condition was also adopted in [7] to regularize subdomain problems. More general local transmission conditions of zero or second order were proposed and analyzed in [10, 11] with parameters optimized for accelerating convergence. More advanced and even non-local transmission conditions can be used, see [3, 12, 18], and also [2, 13] in this volume. In this paper, however, we will restrict ourselves to local transmission conditions.

Another remedy is to modify the usual coarse problem, which probably originated from the multigrid context, first suggested by Achi Brandt and presented in [19]. In their paper [7], Farhat et al. used plane waves on the interface as basis of the coarse space. The idea turns out to be very successful and was followed by Farhat et al. [8], Kimn and Sarkis [15], and Li and Tu [17], and will also be used for the optimized Schwarz methods in this paper. Note that, however, the coarse problem does not change the underlying subdomain problems.

In the following paragraphs, we will give a brief introduction to these methods at the (almost) continuous level.

2.1 The Non-overlapping Methods

We partition the domain into non-overlapping subdomains denoted by $\overline{\Omega} := \cup_i \overline{\Omega_i}$, and we call the set of points shared by more than two subdomains (or shared by two subdomains and the outer boundary $\partial\Omega$) corners. In three dimensions, this includes vertices and edges. We call all the points shared by exactly two subdomains the interface Γ, and in particular a connected component of the interface shared by Ω_i and Ω_j is called interface segment Γ_{ij}.

If we know the Neumann, Dirichlet or Robin data (denoted by λ) of the exact solution on the interface, then we can recover the exact solution from the corresponding boundary value problems defined on subdomains (as long as they are well-posed) with *continuous constraints at corners*. Since on every subdomain there is a recovered solution that gives Dirichlet, Neumann or Robin traces on the interface, we expect for each interface segment Γ_{ij} the traces from Ω_i and Ω_j to be equal. The above process indeed sets up an equation, denoted by $F\lambda = d$, for the interface data λ of the exact solution. For the Helmholtz equation, an additional coarse problem is introduced such that $(I - FQ(Q^*FQ)^{-1}Q^*)F\lambda = (I - FQ(Q^*FQ)^{-1}Q^*)d$ is solved, where the columns of Q are traces of plane waves on the interface.

From the above point of view, we summarize some existing non-overlapping domain decomposition methods in Table 2. The (first-order) absorbing boundary data

is defined as $\lambda := \partial_{\mathbf{n}} u - \mathbf{i}ku$. The lumped preconditioner is the stiffness submatrix $A_{\Gamma\Gamma}$ corresponding to the interface. The first three methods share interface data (up to a sign for the normal derivative) on their common interface segments, and are therefore one-field methods. This is in contrast to the last method, since optimized Schwarz methods have two sets of unknowns on each interface segment, and thus belong to the class of two-field methods. Note also that we do not have suitable preconditioners for the last two methods, which can be a subject for future study.

Table 2. The non-overlapping methods

Algorithms	Unknowns	Matching	Precond.
FETI-DPH ([8])	Neumann	Dirichlet	DtN/lumped
BDDC-H ([17])	Dirichlet	Neumann	NtD
FETI-H ([7])	Absorbing	Dirichlet	(none)
Optimized Schwarz ([10])	two-field Robin	two-field Robin	(none)

2.2 The Overlapping Methods

We partition the domain into overlapping subdomains. We will use the *substructured form*[3] as for the non-overlapping methods in Sect. 2.1. Note that in an overlapping setting, subdomains can not share the same interface data, since the interfaces are in different locations, and therefore all overlapping methods are in some sense two field methods, like the non-overlapping optimized Schwarz methods. The interface data used (both as unknowns and matching conditions) and related references are: Dirichlet [16], absorbing [4, 15], Neumann [14], Robin [9]. A coarse problem as in Sect. 2.1 is adopted but without corner constraints.

3 Numerical Experiments

All the experiments were done in Matlab with sequential codes. We use GMRES with zero initial guess to solve the substructured systems until the relative residual is less than 10^{-6} or the maximum iteration number is attained. The domain is partitioned in a Cartesian way. If we vary the mesh size, then the velocity in (1) is sub-sampled on the coarsest mesh of $24 \times 24 \times 8$.

We introduce the following acronyms:

FL/FD: FETI-DPH with the lumped/DtN preconditioner
FH: FETI-H with corner constraints
O0/O2: non-overlapping optimized Schwarz of zero/second order

[3] Though most of the overlapping methods in the literature are not in this form, we found by numerical experiments it may be cheaper in both time and memory.

OD/ON/OR: overlapping method with Dirichlet/Neumann/absorbing data
OO0/OO2: overlapping optimized Schwarz of zero/second order

For the overlapping methods, the overlapping region has a thickness of two mesh elements and the matching conditions are imposed on faces, edges and vertices, respectively, without repeats on any degrees of freedom. Due to the absence of relevant results, the parameters for the optimized Schwarz methods are not respecting overlapping (except OO0), coarse problem and medium heterogeneity. The plane waves used are along six directions that are normal to the x-y, y-z and z-x planes, respectively.

We found that all the methods outperform the direct solver in CPU time (see Table 1) on the $96 \times 96 \times 32$ mesh. We are interested in how the convergence of these methods depends on the *frequency* f in (1), the *mesh size* h, the *partition* $N_x \times N_y \times N_z$ or the subdomain size H and medium heterogeneity. At $f = 1$ the domain contains nine wavelength along the x-direction, which corresponds to the problem on the unit cube with the wavenumber 18π.

In the following tables, the numbers outside/inside parentheses are the iteration numbers with/without plane waves, respectively, and a bar is used instead of 200 when the maximum iteration number is reached. We use e/w to represent the number of elements per wavelength at the lowest velocity. The smallest iteration numbers among the non-overlapping methods and those among the overlapping methods are in bold. Note that for the FETI-DPH method with DtN preconditioner the amount of work per iteration is about 1.5 times that for the others, and construction of the preconditioner also leads to double LU factorizations in the setup stage.

In Tables 3 and 4, we increase the frequency with fh or f^3h^2 [1] kept constant.

Table 3. Dependence on the frequency (fh =constant)

f	FL	FD	FH	O0	O2	OD	OR	ON	OO0	OO2
partition $3 \times 3 \times 1$										
$\frac{1}{4}$	6 (15)	**4** (8)	9 (15)	15 (21)	8 (14)	8 (20)	8 (12)	9 (20)	7 (15)	**6** (14)
$\frac{1}{2}$	15 (30)	**9** (12)	18 (33)	29 (34)	19 (20)	23 (34)	12 (15)	24 (37)	12 (17)	**11** (13)
1	44 (51)	**20** (23)	75 (93)	43 (48)	25 (25)	51 (58)	17 (17)	57 (66)	22 (25)	**14** (15)
partition scaling with mesh: $H/h = 8$ (see also the first row for $f = \frac{1}{4}$)										
$\frac{1}{2}$	8 (46)	**5** (30)	10 (73)	17 (71)	10 (50)	14 (73)	11 (33)	21 (103)	**8** (55)	**8** (51)
1	9 (183)	**7** (-)	11 (-)	21 (-)	12 (-)	27 (-)	**15** (74)	152 (-)	16 (-)	**15** (-)
partition scaling with mesh: $H/h = 16$ (see also the second row for $f = \frac{1}{2}$)										
1	39 (127)	32 (103)	74 (-)	59 (113)	**27** (39)	76 (171)	26 (38)	114 (-)	26 (53)	**22** (32)

We see that more iterations are usually needed for larger frequency except in the middle of Table 4.

In Table 5, the frequency is fixed and the mesh is refined. From the table, the iteration numbers with plane waves almost remain constant.

Table 4. Dependence on the frequency (f^3h^2=constant)

f	FL	FD	FH	O0	O2	OD	OR	ON	OO0	OO2
partition $3\times3\times1$ (see also the first row in Table 3 for $f=0.25$)										
0.40	12 (25)	**6** (11)	14 (25)	30 (33)	18 (21)	18 (29)	11 (14)	19 (32)	**9** (15)	**9** (13)
0.63	27 (41)	**11** (15)	33 (49)	37 (42)	25 (26)	38 (46)	16 (17)	39 (50)	15 (20)	**13** (14)
partition scaling with mesh: $H/h=8$(see also the first row in Table 4 for f = 0.25)										
0.40	7 (36)	**5** (23)	10 (54)	15 (58)	9 (40)	12 (60)	10 (29)	13 (73)	**7** (40)	**7** (40)
0.63	7 (127)	**5** (100)	9 (149)	14 (156)	8 (112)	14 (160)	11 (65)	20 (-)	**7** (123)	**7** (117)
partition scaling with mesh: $H/h=16$ (see also the first row for $f=0.40$)										
0.63	15 (89)	**8** (53)	18 (119)	43 (125)	18 (75)	33 (113)	16 (35)	36 (112)	**13** (75)	**13** (75)

Table 5. Dependence on the mesh size ($f=\frac{1}{4}$)

e/w	FL	FD	FH	O0	O2	OD	OR	ON	OO0	OO2
partition $3\times3\times1$ (see also the first row in Table 4 for $e/w=10$)										
20	10 (19)	**5** (9)	13 (20)	17 (26)	9 (17)	14 (28)	11 (15)	13 (27)	8 (16)	**6** (16)
40	15 (25)	**6** (10)	18 (25)	21 (32)	11 (20)	21 (39)	15 (19)	19 (36)	9 (17)	**8** (17)
partition $H/h=8$ (see also the first row in Table 4 for $e/w=10$)										
20	7 (21)	**5** (12)	10 (32)	14 (47)	8 (32)	10 (46)	9 (25)	10 (44)	7 (29)	**6** (30)
40	6 (19)	**4** (13)	9 (36)	14 (92)	7 (63)	9 (90)	9 (46)	9 (91)	7 (56)	**6** (59)
partition $H/h=16$ (see also the first row for $e/w=20$)										
40	11 (34)	**6** (15)	14 (47)	17 (60)	10 (38)	15 (63)	12 (28)	13 (52)	**7** (33)	**7** (35)

Next, we compare the iteration numbers for different partitions with both the frequency and the mesh size fixed in Table 6. One can see that with plane waves

Table 6. Dependence on the partition

	FL	FD	FH	O0	O2	OD	OR	ON	OO0	OO2
$\frac{H}{H_0}$	$f=\frac{1}{2}$, mesh and velocity $48\times48\times16$ and H_0 partition $3\times3\times1$									
1	15 (30)	**9** (12)	18 (33)	28 (35)	19 (21)	22 (34)	12 (15)	23 (37)	**11** (17)	**11** (14)
$\frac{1}{2}$	8 (47)	**5** (30)	10 (73)	16 (72)	9 (51)	14 (75)	11 (34)	21 (105)	8 (62)	**7** (57)
$\frac{1}{4}$	**4** (22)	**4** (21)	7 (48)	10 (95)	7 (72)	7 (97)	8 (52)	11 (-)	6 (83)	**5** (78)
	$f=1$, mesh and velocity $96\times96\times32$ and H_0 partition $3\times3\times1$									
1	46 (54)	**22** (24)	79 (97)	45 (49)	26 (26)	54 (61)	17 (18)	60 (69)	22 (26)	**15** (16)
$\frac{1}{2}$	43 (133)	35 (109)	82 (-)	63 (117)	**28** (40)	82 (176)	27 (39)	136 (-)	28 (56)	**24** (34)
$\frac{1}{4}$	10 (184)	**8** (-)	14 (-)	26 (-)	16 (40)	32 (-)	**17** (-)	- (-)	25 (-)	22 (-)
N_x	$f=1$, mesh and velocity $96\times96\times32$ and partition $N_x\times1\times1$									
8	117 (125)	79 (75)	171 (184)	66 (70)	**28** (28)	94 (99)	**23** (24)	100 (104)	51 (46)	**23** (25)
16	184 (-)	192 (199)	- (-)	131 (137)	**45** (47)	- (-)	46 (47)	- (-)	72 (81)	**43** (45)
32	- (-)	- (-)	- (-)	172 (173)	**87** (90)	- (-)	86 (90)	182 (88)	148 (136)	**84** (87)

using more subdomains can both increase and decrease the iteration numbers. It is interesting that for the strip-wise partition only the methods based on transmission conditions (O0, O2, OR, OO0 and OO2) work reliably, though with substantial iteration numbers, and the plane waves do not help much.

Last, we study the influence of the heterogeneity in the velocity. The experiments are carried out on artificial velocity models to have high contrasts. The frequency is fixed as $f = \frac{1}{2}$. The lowest velocity is fixed as $c_{\min} = 1{,}500$ and different levels of highest velocity $c_{\max} = \rho c_{\min}$ are considered. It can be seen from Table 7 that the iteration numbers vary only little.

Table 7. Influence of medium heterogeneity

ρ	FL	FD	FH	O0	O2	OD	OR	ON	OO0	OO2
mesh $48 \times 48 \times 16$, partition $8 \times 1 \times 1$ and $c = c_{\min}, c_{\max}$ on subdomains										
1	58 (76)	37 (46)	83 (94)	60 (64)	**28** (29)	70 (81)	27 (26)	69 (79)	37 (44)	**24** (24)
10^2	28 (36)	42 (58)	30 (37)	37 (55)	**26** (31)	37 (53)	27 (29)	63 (75)	15 (26)	**13** (22)
10^4	32 (36)	49 (58)	33 (37)	45 (55)	**26** (31)	43 (53)	29 (30)	71 (75)	19 (26)	**17** (22)
as above except partition $6 \times 6 \times 2$										
1	9 (90)	**7** (62)	12 (124)	26 (79)	15 (39)	18 (97)	14 (35)	22 (117)	**10** (46)	12 (34)
10^2	12 (59)	**10** (104)	17 (51)	25 (78)	15 (46)	17 (67)	12 (34)	29 (100)	**8** (42)	9 (37)
10^4	14 (58)	**11** (104)	19 (51)	27 (79)	17 (47)	19 (68)	12 (34)	33 (100)	**8** (42)	10 (37)
mesh $48 \times 48 \times 16$, partition $1 \times 8 \times 1$ and $c = c_{\min}, c_{\max}$ on $8 \times 1 \times 1$ cells										
1	70 (81)	40 (50)	105 (114)	73 (75)	**27** (28)	74 (80)	28 (27)	62 (66)	34 (37)	**24** (24)
10^2	51 (59)	30 (34)	69 (84)	58 (67)	**26** (28)	56 (67)	**23** (26)	51 (59)	26 (28)	**23** (26)
10^4	52 (59)	30 (34)	70 (85)	58 (67)	**26** (28)	56 (68)	**23** (26)	51 (59)	26 (28)	**23** (26)
mesh $84 \times 84 \times 24$, partition $6 \times 6 \times 2$ and $c = c_{\min}, c_{\max}$ on $7 \times 7 \times 3$ cells										
1	12 (105)	**8** (65)	16 (144)	34 (96)	19 (41)	24 (121)	17 (37)	25 (111)	**12** (46)	15 (34)
10^2	10 (68)	**7** (34)	14 (107)	29 (109)	17 (48)	26 (111)	13 (45)	21 (106)	**11** (47)	12 (40)
10^4	11 (68)	**7** (34)	15 (107)	31 (109)	18 (48)	26 (110)	14 (45)	21 (107)	**11** (47)	12 (40)
mesh $48 \times 48 \times 16$, partition $6 \times 6 \times 2$ and c random constants on elements										
10^2	7 (16)	**5** (10)	10 (21)	14 (61)	9 (41)	14 (60)	11 (37)	12 (59)	**7** (35)	8 (38)
10^4	8 (15)	**6** (9)	11 (20)	12 (67)	8 (46)	14 (67)	15 (61)	25 (86)	**8** (39)	**8** (42)
as above except partition $3 \times 3 \times 1$										
1	22 (38)	**10** (16)	26 (45)	28 (37)	19 (21)	26 (36)	13 (15)	27 (36)	15 (21)	**12** (14)
10^2	11 (17)	**6** (8)	15 (20)	18 (33)	11 (21)	16 (35)	15 (23)	16 (42)	**7** (17)	8 (19)
10^4	12 (17)	**6** (8)	16 (21)	15 (39)	9 (24)	18 (40)	16 (31)	17 (52)	**8** (20)	9 (22)

4 Conclusions

For the SEG–SALT model on the cube domain, we get the following conclusions: among the non-overlapping methods, the FETI-DPH method with DtN preconditioner performs best in terms of iteration numbers. Among the overlapping methods,

the optimized Schwarz method of second order is usually the best. With a fixed number of plane waves, all the methods can slow down for larger frequencies on properly refined meshes. They also deteriorate for fixed frequency on finer meshes, unless when using plane waves and more subdomains. A smaller subdomain size can both increase and decrease the iteration numbers, and the experiments indicate the existence of some optimal choice. For strip-wise partitions, only the methods based on transmission conditions work well, and plane waves do not help much. We also find the performance of all the method is only little affected by the heterogeneity in the velocity we considered, but other kinds of heterogeneity still need to be investigated.

Acknowledgments The authors thank Paul Childs for providing the velocity data of the geophysical SEG–SALT model. This work was partially supported by the University of Geneva. The second author was also partially supported by the NSFC Tianyuan Mathematics Youth Fund 10926134.

Bibliography

[1] Ivo Babuska, Frank Ihlenburg, Ellen T. Paik, and Stefan A. Sauter. A generalized finite element method for solving the Helmholtz equation in two dimensions with minimal pollution. *Comput. Methods Appl. Mech. Engrg.*, 128(3-4): 325–359, 1995.

[2] Yassine Boubendir, Xavier Antoine, and Christophe Geuzaine. A non-overlapping quasi-optimal optimized Schwarz domain decomposition algorithm for the Helmholtz equation. In *Domain Decomposition Methods in Science and Engineering XX*, pages 547–554, 2013.

[3] Yassine Boubendir, Xavier Antoine, and Christophe Geuzaine. A quasi-optimal non-overlapping domain decomposition algorithm for the Helmholtz equation. *J. Comput. Phys.*, 231(2):262–280, 2012.

[4] Xiao-Chuan Cai, Mario A. Casarin, Frank W. Elliott, Jr., and Olof B. Widlund. Overlapping Schwarz algorithms for solving Helmholtz's equation. In *Domain decomposition methods, 10 (Boulder, CO, 1997)*, volume 218 of *Contemp. Math.*, pages 391–399. Amer. Math. Soc., Providence, RI, 1998.

[5] Bruno Després. Domain decomposition method and the Helmholtz problem. In *Mathematical and numerical aspects of wave propagation phenomena (Strasbourg, 1991)*, pages 44–52. SIAM, Philadelphia, PA, 1991.

[6] Olivier G. Ernst and Martin J. Gander. Why it is difficult to solve Helmholtz problems with classical iterative methods. In *Numerical Analysis of Multiscale Problems*. Durham LMS Symposium 2010, Springer Verlag, 2011.

[7] Charbel Farhat, Antonini Macedo, and Michel Lesoinne. A two-level domain decomposition method for the iterative solution of high frequency exterior Helmholtz problems. *Numer. Math.*, 85(2):283–308, 2000.

[8] Charbel Farhat, Philip Avery, Radek Tezaur, and Jing Li. FETI-DPH: a dual-primal domain decomposition method for acoustic scattering. *J. Comput. Acoust.*, 13(3):499–524, 2005.

[9] Martin J. Gander. Optimized Schwarz methods for Helmholtz problems. In *Domain decomposition methods in science and engineering (Lyon, 2000)*, Theory Eng. Appl. Comput. Methods, pages 247–254. Internat. Center Numer. Methods Eng. (CIMNE), Barcelona, 2002.

[10] Martin J. Gander, Frédéric Magoulès, and Frédéric Nataf. Optimized Schwarz methods without overlap for the Helmholtz equation. *SIAM J. Sci. Comput.*, 24 (1):38–60 (electronic), 2002.

[11] Martin J. Gander, Laurence Halpern, and Frédéric Magoulés. An optimized Schwarz method with two-sided Robin transmission conditions for the Helmholtz equation. *Int. J. Numer. Meth. Fluids*, 55(2):163–175, 2007.

[12] Souad Ghanemi. A domain decomposition method for Helmholtz scattering problems. In *Ninth International Conference on Domain Decomposition Methods*, pages 105–112, 1998.

[13] Murthy N. Guddati and Senganal Thirunavukkarasu. Improving the convergence of Schwarz methods for Helmholtz equation. In *Domain Decomposition Methods in Science and Engineering XX*, pages 207–214, 2013.

[14] Jung-Han Kimn and Blaise Bourdin. Numerical implementation of overlapping balancing domain decomposition methods on unstructured meshes. In *Domain decomposition methods in science and engineering XVI*, volume 55 of *Lect. Notes Comput. Sci. Eng.*, pages 309–315. Springer, Berlin, 2007.

[15] Jung-Han Kimn and Marcus Sarkis. Restricted overlapping balancing domain decomposition methods and restricted coarse problems for the Helmholtz problem. *Comput. Methods Appl. Mech. Engrg.*, 196(8):1507–1514, 2007.

[16] An Leong and Olof B. Widlund. Extension of two-level Schwarz preconditioners to symmetric indefinite problems. Technical report, New York University, New York, NY, USA, 2008.

[17] Jing Li and Xuemin Tu. Convergence analysis of a balancing domain decomposition method for solving a class of indefinite linear systems. *Numer. Linear Algebra Appl.*, 16(9):745–773, 2009.

[18] Bruno Stupfel. Improved transmission conditions for a one-dimensional domain decomposition method applied to the solution of the Helmholtz equation. *J. Comput. Phys.*, 229(3):851–874, 2010.

[19] Shlomo Ta'asan. *Multigrid methods for highly oscillatory problems*. PhD thesis, Weizmann Institute of Science, Rehovot, Israel, 1984.

Stable BETI Methods in Electromagnetics

Olaf Steinbach and Markus Windisch

Institute of Computational Mathematics, TU Graz, Steyrergasse 30, 8010 Graz, Austria,
o.steinbach@tugraz.at, mwind82@gmx.at

Summary. In this paper we present a stable boundary element tearing and interconnecting domain decomposition method for the parallel solution of the electromagnetic wave equation with piecewise constant wave numbers. In particular we consider stable boundary integral formulations and generalized Robin type transmission conditions to ensure unique solvability of the local subproblems. Numerical results confirm the robustness of the proposed approach.

1 Introduction

The application of standard finite and boundary element tearing and interconnecting domain decomposition methods [4, 5] may fail in the case of the acoustic or electromagnetic wave equation due to a possible occurence of spurious modes which are related to local Dirichlet or Neumann boundary value problems. For the acoustic wave equation we have introduced in [9, 10] a boundary element tearing and interconnecting domain decomposition approach which is stable for all local wave numbers. The aim of this paper is to extend these results when considering the electromagnetic wave equation. Although the general concept is rather similar in both cases, the numerical analysis of boundary integral equations and boundary element methods for the Maxwell system requires advanced techniques, in particular appropriate space splitting approaches. For the definition of Sobolev spaces which are related to the Maxwell equation, see, e.g., [2], for the analysis of Maxwell boundary integral equations, see, for example, [7], and for related boundary element methods, see, e.g., [1].

2 Formulation of the Domain Decomposition Approach

As a model problem we consider the Neumann boundary value problem of the electromagnetic wave equation

$$\mathbf{curl}\,\mathbf{curl}\,\mathbf{U}(x) - [k(x)]^2\mathbf{U}(x) = \mathbf{0} \qquad \text{for } x \in \Omega, \tag{1}$$

$$\gamma_N \mathbf{U}(x) := \mathbf{curl}\,\mathbf{U}(x) \times \mathbf{n} = \mathbf{f}(x) \quad \text{for } x \in \Gamma, \tag{2}$$

R. Bank et al. (eds.), *Domain Decomposition Methods in Science and Engineering XX*, Lecture Notes in Computational Science and Engineering 91,
DOI 10.1007/978-3-642-35275-1_25,

where $\Omega \subset \mathbb{R}^3$ is a Lipschitz polyhedron with boundary $\Gamma = \partial\Omega$. We assume that the boundary value problems (1) and (2) admits a unique solution. Since the wave number $k(x)$ is assumed to be piecewise constant, i.e. $k(x) = k_i$ for $x \in \Omega_i$, instead of (1) and (2) we consider local boundary value problems to find $\mathbf{U}_i = \mathbf{U}_{|\Omega_i}$ satisfying

$$\mathbf{curl}\,\mathbf{curl}\,\mathbf{U}_i(x) - k_i^2\mathbf{U}_i(x) = \mathbf{0} \text{ for } x \in \Omega_i,\ \gamma_N\mathbf{U}_i(x) = \mathbf{g}(x) \text{ for } x \in \Gamma_i \cap \Gamma$$

with respect to a non–overlapping domain decomposition

$$\overline{\Omega} = \bigcup_{i=1}^{p} \overline{\Omega}_i, \quad \Omega_i \cap \Omega_j = \emptyset \quad \text{for } i \neq j, \quad \Gamma_i = \partial\Omega_i,$$

together with the transmission or interface boundary conditions

$$\gamma_{D,i}\mathbf{U}_i(x) = \gamma_{D,j}\mathbf{U}_j(x) \quad \text{for } x \in \Gamma_{ij} = \Gamma_i \cap \Gamma_j, \tag{3}$$

$$\gamma_{N,i}\mathbf{U}_i(x) + \gamma_{N,j}\mathbf{U}_j(x) = \mathbf{0} \quad \text{for } x \in \Gamma_{ij}, \tag{4}$$

where the Dirichlet trace operator is given by

$$\gamma_D\mathbf{U} = \mathbf{n} \times (\mathbf{U}_{|\Gamma} \times \mathbf{n}).$$

Since the local Dirichlet or Neumann boundary value problems may exhibit spurious modes, instead of the Neumann transmission condition in (4) we consider a generalized Robin interface condition

$$\gamma_{N,i}\mathbf{U}_i(x) + \gamma_{N,j}\mathbf{U}_j(x) + i\eta_{ij}\mathsf{R}_{ij}[\gamma_{D,i}\mathbf{U}_i(x) - \gamma_{D,j}\mathbf{U}_j(x)] = \mathbf{0} \quad \text{for } x \in \Gamma_{ij}, i < j. \tag{5}$$

The operators R_{ij} are assumed to be strictly positive, i.e. $\langle \mathsf{R}_{ij}\mathbf{u}, \mathbf{u}\rangle_{\Gamma_{ij}} > 0$ for all $\mathbf{u} \in \mathbf{H}_{\perp}^{-1/2}(\mathrm{curl}_{\Gamma}, \Gamma_{ij})$, and $\eta_{ij} \in \mathbb{R}\backslash\{0\}$. We define

$$(\mathsf{R}_i u_{|\Gamma_i})(x) := (\mathsf{R}_{ij} u_{|\Gamma_{ij}})(x) \quad \text{for } x \in \Gamma_{ij}$$

and

$$\eta_i(x) := \begin{cases} \eta_{ij} & \text{for } x \in \Gamma_{ij},\ i < j, \\ -\eta_{ij} & \text{for } x \in \Gamma_{ij},\ i > j, \\ 0 & \text{for } x \in \Gamma_i \cap \Gamma, \end{cases}$$

where we assume that $\eta_i(x)$ for $x \in \Gamma_i$ does not change its sign, see also [9]. In this case we can ensure unique solvability [11] of the local Robin boundary value problems

$$\mathbf{curl}\,\mathbf{curl}\,\mathbf{U}_i(x) - k_i^2\mathbf{U}_i(x) = \mathbf{0} \quad \text{for } x \in \Omega_i, \tag{6}$$

$$\gamma_N\mathbf{U}_i(x) + i\eta_i\mathsf{R}\gamma_D\mathbf{U}_i(x) = \mathbf{g}(x) \quad \text{for } x \in \Gamma_i \cap \Gamma. \tag{7}$$

For the solution of local Dirichlet and Robin boundary value problems we will apply boundary element methods which are based on the use of the Stratton–Chu representation formula for $x \in \Omega$, see [3],

$$\mathbf{U}(x) = \Psi_k^M(\gamma_D\mathbf{U})(x) + \Psi_k^A(\gamma_N\mathbf{U})(x) + \frac{1}{k^2}\mathbf{grad}\,\Psi_k^S \mathrm{div}_\Gamma(\gamma_N\mathbf{U})(x).$$

Here,

$$\Psi_k^A(\lambda)(x) := \int_\Gamma g_k(x,y)\lambda(y)ds_y \quad \text{for } x \notin \Gamma, \quad g_k(x,y) = \frac{1}{4\pi}\frac{e^{ik|x-y|}}{|x-y|},$$

is the vector–valued single layer potential with the fundamental solution of the Helmholtz equation, and

$$\Psi_k^M(\lambda)(x) := \mathbf{curl}\,\Psi_k^A(\lambda \times \mathbf{n})(x) \quad \text{for } x \notin \Gamma$$

is the Maxwell double layer potential. In addition,

$$\Psi_k^V(\lambda)(x) := \int_\Gamma g_k(x,y)\lambda(y)ds_y \quad \text{for } x \notin \Gamma$$

is the scalar single layer potential. By introducing the Maxwell single layer potential

$$\Psi_k^S(\lambda)(x) := \Psi_k^A(\lambda)(x) + \frac{1}{k^2}\mathbf{grad}\,\Psi_k^S \mathrm{div}_\Gamma(\lambda)(x) \quad \text{for } x \notin \Gamma,$$

we can write the Straton–Chu representation formula as

$$\mathbf{U}(x) = \Psi_k^M(\gamma_D\mathbf{U}(x)) + \Psi_k^S(\gamma_N\mathbf{U}(x)) \quad \text{for } x \in \Omega. \tag{8}$$

The application of the Maxwell trace operators gives the boundary integral equations [7, 11]

$$\begin{aligned} \gamma_N\mathbf{U} &= \mathsf{N}_k(\gamma_D\mathbf{U}) + (\frac{1}{2}I + \mathsf{B}_k)(\gamma_N\mathbf{U}), \\ \gamma_D\mathbf{U} &= (\frac{1}{2}I + \mathsf{C}_k)(\gamma_D\mathbf{U}) + \mathsf{S}_k(\gamma_N\mathbf{U}). \end{aligned} \tag{9}$$

Now we are in a position to derive different approaches to solve local boundary value problems with generalized Robin boundary conditions. Here we consider an approach which is based on the use of the Steklov–Poincaré operator

$$\mathsf{T}_k = \mathsf{N} + (\frac{1}{2}I + \mathsf{B}_k)\mathsf{S}_k^{-1}(\frac{1}{2}I + \mathsf{C}_k) = \mathsf{S}_k^{-1}(\frac{1}{2}I + \mathsf{C}_k) \tag{10}$$

which requires the invertibility of the single layer operator S_k. Since S_k is not invertible for all wave numbers k, instead of (10) we consider a system of boundary integral equations to find $\mathbf{u} \in \mathbf{H}_\parallel^{-1/2}(\mathrm{div}_\Gamma, \Gamma)$ and $\mathbf{t} \in \mathbf{H}_\perp^{-1/2}(\mathrm{curl}_\Gamma, \Gamma)$ such that

$$\begin{pmatrix} \mathsf{N}_k + i\eta\mathsf{R} & \frac{1}{2}I + \mathsf{B}_k \\ -\frac{1}{2}I + \mathsf{C}_k & \mathsf{S}_k \end{pmatrix} \begin{pmatrix} \mathbf{u} \\ \mathbf{t} \end{pmatrix} = \begin{pmatrix} \mathbf{g} \\ \mathbf{0} \end{pmatrix} \tag{11}$$

is satisfied. The unique solvability of (11) follows from a generalized Garding inequality

$$\mathrm{Re}\left(\left\langle\begin{pmatrix}\mathsf{N}_k+i\eta\mathsf{R} & \frac{1}{2}I+\mathsf{B}_k\\ -\frac{1}{2}I+\mathsf{C}_k & \mathsf{S}_k\end{pmatrix}\begin{pmatrix}\mathbf{u}\\ \mathbf{t}\end{pmatrix},\begin{pmatrix}\mathscr{Y}\mathbf{u}\\ \mathscr{X}\mathbf{t}\end{pmatrix}\right\rangle_{\Gamma}+C((\mathbf{u},\mathbf{t}),(\mathbf{u},\mathbf{t}))\right)$$
$$\geq c\left(\|\mathbf{u}\|^2_{\mathbf{H}_{\perp}^{-1/2}(\mathrm{curl}_{\Gamma},\Gamma)}+\|\mathbf{t}\|^2_{\mathbf{H}_{\parallel}^{-1/2}(\mathrm{div}_{\Gamma},\Gamma)}\right)$$

for some appropriate bijective operators $\mathscr{X}$ and $\mathscr{Y}$, and from injectivity which is in fact related to the unique solvability of the local Robin boundary value problems (6) and (7), see [11]. Since the proof of the generalized Garding inequality requires a comprehensive study of the trace spaces $\mathbf{H}_{\perp}^{-1/2}(\mathrm{curl}_{\Gamma},\Gamma)$ and $\mathbf{H}_{\parallel}^{-1/2}(\mathrm{div}_{\Gamma},\Gamma)$, and of the corresponding Hodge–type splittings, we refer to [2, 11] for a detailed presentation.

By summing up all local boundary integral equation systems with respect to the transmission conditions (5) we finally obtain the following variational formulation to find $\mathbf{u}\in\mathbf{H}_{\perp}^{-1/2}(\mathrm{curl}_{\Gamma},\Gamma_S)$ and $\mathbf{t}_i\in\mathbf{H}_{\parallel}^{-1/2}(\mathrm{div}_{\Gamma},\Gamma_i)$ satisfying

$$\sum_{i=1}^{p}\left[\langle\mathsf{N}_i\mathbf{u}_{|\Gamma_i},\mathbf{v}_{|\Gamma_i}\rangle_{\Gamma_i}+\langle(\frac{1}{2}I+\mathsf{B}_i)\mathbf{t}_i,\mathbf{v}_{|\Gamma_i}\rangle_{\Gamma_i}+i\eta_i\langle\mathsf{R}_i\mathbf{u}_{|\Gamma_i},\mathbf{v}_{|\Gamma_i}\rangle_{\Gamma_i}\right]=\langle\mathbf{f},\mathbf{v}\rangle_{\Gamma}\tag{12}$$

for all $\mathbf{v}\in\mathbf{H}_{\parallel}^{-1/2}(\mathrm{div}_{\Gamma},\Gamma_S)$ and

$$\langle\mathsf{S}_i\mathbf{t}_i,\mu_i\rangle_{\Gamma_i}+\langle(-\frac{1}{2}I+\mathsf{C}_i)\mathbf{u}_{|\Gamma_i},\mu_i\rangle_{\Gamma_i}=0\tag{13}$$

for all $\mu_i\in\mathbf{H}_{\parallel}^{-1/2}(\mathrm{div}_{\Gamma},\Gamma_i), i=1,\dots,p$. The variational formulation (12), (13) admits a unique solution iff the orginal problems (1) and (2) has a unique solution, see [11].

A boundary element discretization of the Sobolev spaces $\mathbf{H}_{\perp}^{-1/2}(\mathrm{curl}_{\Gamma},\Gamma_S)$ and $\mathbf{H}_{\parallel}^{-1/2}(\mathrm{div}_{\Gamma},\Gamma_i)$ by using Raviart–Thomas elements [8, 11], i.e.

$$\mathscr{E}_h:=\mathscr{E}_h(\Gamma_S)=\mathrm{span}\{\phi_k\}_{k=1}^{M_S}\subset\mathbf{H}_{\perp}^{-1/2}(\mathrm{curl}_{\Gamma},\Gamma_S)$$

and

$$\mathscr{F}_{i,h}=\mathrm{span}\{\psi_k^i\}_{k=1}^{N_i}\subset\mathbf{H}_{\parallel}^{-1/2}(\mathrm{div}_{\Gamma},\Gamma_i),$$

then results in a linear system of algebraic equations,

$$\begin{pmatrix}\mathsf{S}_{1,h} & & & \widetilde{\mathsf{C}}_{1,h}A_i\\ & \cdots & & \vdots\\ & & \mathsf{S}_{p,h} & \widetilde{\mathsf{C}}_{p,h}A_p\\ A_1^{\top}\widetilde{\mathsf{B}}_{1,h} & \dots & A_p^{\top}\widetilde{\mathsf{B}}_{p,h} & \sum_{i=1}^{p}A_i^{\top}[\mathsf{N}_{i,h}+i\eta_i\mathsf{R}_{i,h}]A_i\end{pmatrix}\begin{pmatrix}\underline{t}_1\\ \vdots\\ \underline{t}_p\\ \underline{u}\end{pmatrix}=\begin{pmatrix}\underline{0}\\ \vdots\\ \underline{0}\\ \sum_{i=1}^{p}A_i^{\top}\underline{f}_i\end{pmatrix},\tag{14}$$

where the block matrices are given by

$$\begin{aligned}
\mathsf{S}_{i,h}[\ell,k] &= \langle \mathsf{S}_i \psi_k^i, \psi_\ell^i \rangle_{\Gamma_i}, \\
\widetilde{\mathsf{C}}_{i,h}[\ell,n] &= \langle (-\frac{1}{2} I + \mathsf{C}_i)\phi_n^i, \psi_\ell^i \rangle_{\Gamma_i}, \\
\widetilde{\mathsf{B}}_{i,h}[m,k] &= \langle (\frac{1}{2} I + \mathsf{B}_i)\psi_k^i, \phi_m^i \rangle_{\Gamma_i}, \\
\mathsf{N}_{i,h}[m,n] &= \langle \mathsf{N}_i \phi_n^i, \phi_m^i \rangle_{\Gamma_i}, \\
\mathsf{R}_{i,h}[m,n] &= \langle \mathsf{R}_i \phi_n^i, \phi_m^i \rangle_{\Gamma_i}
\end{aligned}$$

for $k,\ell = 1,\ldots,N_i$, $m,n = 1,\ldots,M_i$, and $i = 1,\ldots,p$.

In what follows we will discuss an efficient and parallel solution of the linear system (14). Although the computation of all block matrices can be done in parallel, the construction of an appropriate preconditioner is more challenging. A possible approach is to design preconditioners as in tearing and interconnecting methods which are well established for a wide range of applications. A first step into this direction is the formulation of stable tearing and interconnecting methods.

The idea of the tearing and interconnecting approach is to tear the global degrees of freedom, which are given by $\underline{u}$, into local degrees of freedom $\underline{u}_i$. To ensure global continuity, we need to glue them together by using Lagrange multipliers [10, 11], see also Fig. 1. Note, that instead of Neumann transmission condition we use the generalized Robin transmission conditions as given in (5). As in the standard tearing and interconnecting approach this leads to the extended linear system

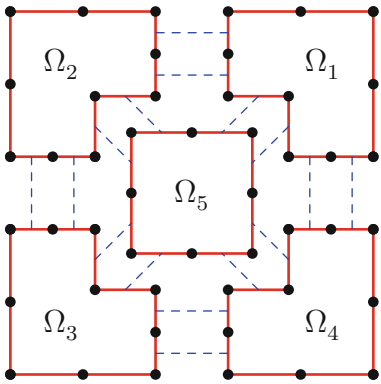


Fig. 1. Tearing and Interconnecting for edge based trial functions

$$\begin{pmatrix} \mathsf{N}_{1,h}+i\eta_1\mathsf{R}_{i,h} & \widetilde{\mathsf{B}}_{1,h} & & & -B_1^\top \\ \widetilde{\mathsf{C}}_{1,h} & \mathsf{S}_{1,h} & & & \\ & & \ddots & & \vdots \\ & & & \mathsf{N}_{p,h}+i\eta_p\mathsf{R}_{p,h} & \widetilde{\mathsf{B}}_{p,h} & -B_p^\top \\ & & & \widetilde{\mathsf{C}}_{p,h} & \mathsf{S}_{p,h} & \\ B_1 & & \dots & B_p & & \end{pmatrix} \begin{pmatrix} \underline{u}_1 \\ \underline{t}_1 \\ \vdots \\ \underline{u}_p \\ \underline{t}_p \\ \underline{\lambda} \end{pmatrix} = \begin{pmatrix} \underline{f}_1 \\ \underline{0} \\ \vdots \\ \underline{f}_p \\ \underline{0} \\ \underline{0} \end{pmatrix} \tag{15}$$

where the sparse and Boolean matrices B_i ensure the continuity of the global solution. Since the local Robin boundary value problems (6) and (7) are uniquely solvable, the local block matrices are invertible, and we can consider the Schur complement system

$$\begin{aligned} \sum_{i=1}^{p} (0\ B_i) \begin{pmatrix} \mathsf{N}_{i,h}+i\eta_i\mathsf{R}_{i,h} & \widetilde{\mathsf{B}}_{i,h} \\ \widetilde{\mathsf{C}}_{i,h} & \mathsf{S}_{i,h} \end{pmatrix}^{-1} \begin{pmatrix} B_i^\top \underline{\lambda} \\ \underline{0} \end{pmatrix} \\ = -\sum_{i=1}^{p} (B_i\ 0) \begin{pmatrix} \mathsf{N}_{i,h}+i\eta_i\mathsf{R}_{i,h} & \widetilde{\mathsf{B}}_{i,h} \\ \widetilde{\mathsf{C}}_{i,h} & \mathsf{S}_{i,h} \end{pmatrix}^{-1} \begin{pmatrix} \underline{f}_i \\ \underline{0} \end{pmatrix}. \end{aligned} \tag{16}$$

Note that (16) corresponds to the adjoint system of standard tearing and interconnecting approaches [4, 5].

3 Numerical Results

As a first example we consider the Neumann boundary value problem

$$\begin{aligned} \mathbf{curl}\,\mathbf{curl}\,\mathbf{U} - k^2\mathbf{U} &= \mathbf{0} \qquad \text{in } \Omega, \\ \gamma_N\mathbf{U} &= \mathbf{f} \qquad \text{on } \Gamma \end{aligned} \tag{17}$$

where the domain Ω is given by $(-1.0, 1.5) \times (0.0, 1.0) \times (0.0, 1.0)$, and Ω is divided into two subdomains Ω_i by the yz–plane, see Fig. 2.

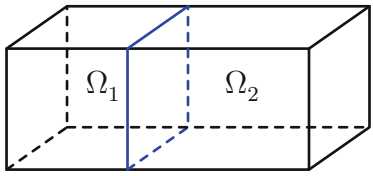


Fig. 2. Computational domain Ω and domain decomposition

As an analytical solution for both examples we use

$$\mathbf{U}(x) = \left[\frac{1+ikr-k^2r^2}{r^3} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} - \frac{3+3ikr-k^2r^2}{r^5} (x_1-\hat{x}_1) \begin{pmatrix} x_1-\hat{x}_1 \\ x_2-\hat{x}_2 \\ x_3-\hat{x}_3 \end{pmatrix} \right] e^{ikr}$$

with $r = |x - \hat{x}|$ and $\hat{x} = (-3.0, 2.1, 1.1)^\top$. The boundary element discretization of the coupled variational formulation (12) and (13) is done with respect to a globally uniform boundary element mesh with E_i edges per subdomain Ω_i, and by using first order Raviart–Thomas elements. The number of Lagrange multipliers is denoted by Λ. The linear system (16) is solved by a GMRES method with a relative residuum reduction of $\varepsilon = 10^{-7}$. For our numerical tests we consider two different wave numbers: The first one is $k = 1.0$ and the second one is the first Dirichlet and Neumann eigenfrequency of the unit cube Ω_1, $k = \sqrt{2}\pi \approx 4.44288$. The results are given in Table 1, where the error is the relative $L_2(\Gamma_1)$ error of the lowest order Raviart–Thomas approximation of the local Dirichlet datum $\mathbf{u}_1$.

E_i	Λ	iter	error
36	8	5	0.1824189
144	28	17	0.0895037
576	104	49	0.0440296
2304	400	142	0.0234164

E_i	Λ	iter	error
36	8	5	0.7042192
144	28	19	0.3055468
576	104	47	0.1472184
2304	400	104	0.0772003

Table 1. Iteration numbers and errors for $k = 1$ (left) and $k = \sqrt{2}\pi$ (right).

In a second example we consider the Neumann boundary value problem (17) for the unit cube $\Omega = (0,1)^3$ which is divided into eight subcubes Ω_i. The results for two different wave numbers $k = 1.0, 8.0$ are given in Table 2.

E_i	Λ	iter	error
36	90	60	0.1133393
144	324	147	0.0550944
576	1224	476	0.0266769

E_i	Λ	iter	error
36	90	60	0.9432815
144	324	153	0.3776120
576	1224	397	0.1769975

Table 2. Iteration numbers and errors for $k = 1$ (left) and $k = 8$ (right).

Both numerical experiments confirm the stability and robustness of the proposed approach, and the theoretical error estimate as given in [11], i.e. we expect a linear order of convergence when using lowest order Raviart–Thomas elements. Note that the linear system (16) is solved by a GMRES method without preconditioner. Hence we observe a rapidly increasing number of required iterations. Therefore, the use of local and global preconditioners is mandatory for the solution of problems of practical interest. Probably, possible preconditioners can be constructed as in the acoustic scattering case, see [11]. Another possibility is to consider a dual–primal approach as in [6].

Acknowledgments This work was supported by the Austrian Science Fund (FWF) within the project *Data sparse boundary and finite element domain decomposition methods in electromagnetics* under the grant P19255.

Bibliography

[1] A. Buffa, R. Hiptmair, T. von Petersdorff, and C. Schwab. Boundary element methods for Maxwell transmission problems in Lipschitz domains. *Numer. Math.*, 95:459–485, 2003.

[2] A. Buffa and P. Ciarlet Jr. On traces for functional spaces related to Maxwell's equations. I. An integration by parts formula in Lipschitz polyhedra. *Math. Methods Appl. Sci.*, 24:9–30, 2001.

[3] L. J. Chu and J. A. Stratton. Diffraction theory of electromagnetic waves. *Phys. Rev.*, 56:99–107, 1939.

[4] C. Farhat and F.-X. Roux. A method of finite element tearing and interconnecting and its parallel solution algorithm. *Int. J. Numer. Meth. Engrg.*, 32:1205–1227, 1991.

[5] U. Langer and O. Steinbach. Boundary element tearing and interconnecting methods. *Computing*, 71:205–228, 2003.

[6] Y. Li and J.-M. Jin. A vector dual-primal finite element tearing and interconnecting method for solving 3-D large-scale electromagnetic problems. *IEEE Trans. Antennas Propag.*, 54:3000–3009, 2006.

[7] J.-C. Nédélec. *Acoustic and electromagnetic equations*, volume 144 of *Applied Mathematical Sciences*. Springer, New York, 2001.

[8] P.-A. Raviart and J. M. Thomas. A mixed finite element method for 2nd order elliptic problems. In *Mathematical aspects of finite element methods*, volume 606 of *Lecture Notes in Mathematics*, pages 292–315. Springer, Berlin, 1977.

[9] O. Steinbach and M. Windisch. Robust boundary element domain decomposition solvers in acoustics. In Y. Huang et al., editor, *Domain Decomposition Methods in Science and Engineering XIX*, volume 78 of *Lecture Notes in Computational Science and Engineering*, pages 277–284. Springer, Berlin, Heidelberg, 2011.

[10] O. Steinbach and M. Windisch. Stable boundary element domain decomposition methods for the Helmholtz equation. *Numer. Math.*, 118:171–195, 2011.

[11] M. Windisch. *Boundary element tearing and interconnecting methods for acoustic and electromagnetic scattering*. PhD thesis, Institute of Computational Mathematics, TU Graz, 2010.

Preconditioning High–Order Discontinuous Galerkin Discretizations of Elliptic Problems

Paola F. Antonietti[1] and Paul Houston[2*]

[1] MOX, Dipartimento di Matematica, Politecnico di Milano, Piazza Leonardo da Vinci 32, 20133 Milano ITALY. paola.antonietti@polimi.it

[2] School of Mathematical Sciences, University of Nottingham, University Park, Nottingham, NG7 2RD, UK. Paul.Houston@nottingham.ac.uk

1 Introduction

In recent years, attention has been devoted to the development of efficient iterative solvers for the solution of the linear system of equations arising from the discontinuous Galerkin (DG) discretization of a range of model problems. In the framework of two level preconditioners, scalable non-overlapping Schwarz methods have been proposed and analyzed for the h–version of the DG method in the articles [1, 2, 6, 7, 9]. Recently, in [3] it has been proved that the non-overlapping Schwarz preconditioners can also be successfully employed to reduce the condition number of the stiffness matrices arising from a wide class of high–order DG discretizations of elliptic problems. In this article we aim to validate the theoretical results derived in [3] for the multiplicative Schwarz preconditioner and for its symmetrized variant by testing their numerical performance.

2 Model Problem and DG Discretization

In this section we introduce the model problem under consideration and its DG approximation, working, for the sake of simplicity, with the SIPG formulation proposed in [4].

We consider, for simplicity, the weak formulation of the Poisson problem with homogeneous Dirichlet boundary conditions: find $\mathscr{U} \in H_0^1(\Omega)$ such that

$$(\nabla \mathscr{U}, \nabla v)_\Omega = (f, v)_\Omega \quad \forall v \in H_0^1(\Omega), \tag{1}$$

where Ω is a bounded polygonal domain in $\mathbb{R}^d$, $d = 2, 3$, $f \in L^2(\Omega)$ is a given source term and $(\cdot,\cdot)_\Omega$ is the standard inner product in $[L^2(\Omega)]^d$.

* PH acknowledges the financial support of the EPSRC under the grant EP/H005498.

R. Bank et al. (eds.), *Domain Decomposition Methods in Science and Engineering XX*, Lecture Notes in Computational Science and Engineering 91,
DOI 10.1007/978-3-642-35275-1_26,

Let $\mathcal{T}_h$ be a shape-regular, not necessarily matching partition of Ω into disjoint open elements $\mathcal{K}$ (with diameter $h_\mathcal{K}$), where each $\mathcal{K}$ is the affine image of a fixed master element $\widehat{\mathcal{K}}$, i.e., $\mathcal{K} = F_\mathcal{K}(\widehat{\mathcal{K}})$, where $\widehat{\mathcal{K}}$ is either the open unit d-simplex or the d-hypercube in $\mathbb{R}^d$, $d = 2, 3$. We define the mesh-size h by $h := \max_{\mathcal{K}\in\mathcal{T}_h} h_\mathcal{K}$, and assume that $\mathcal{T}_h$ satisfies a *bounded local variation* property: for any pair of neighboring elements $\mathcal{K}_1, \mathcal{K}_2 \in \mathcal{T}_h$, $h_{\mathcal{K}_1} \approx h_{\mathcal{K}_2}$.

For a given approximation order $p \geq 1$, we define the DG space

$$V_{h,p} := \{v \in L^2(\Omega) : v|_\mathcal{K} \circ F_\mathcal{K} \in \mathcal{M}^p(\widehat{\mathcal{K}})\ \forall \mathcal{K} \in \mathcal{T}_h\},$$

where $\mathcal{M}^p(\widehat{\mathcal{K}})$ is either the space of polynomials of degree at most p on $\widehat{\mathcal{K}}$, if $\widehat{\mathcal{K}}$ is the reference d-simplex, or the space of polynomials of degree at most p in each variable on $\widehat{\mathcal{K}}$, if $\widehat{\mathcal{K}}$ is the reference d-hypercube.

Next, for any internal face $\overline{F} = \overline{\partial\mathcal{K}^+} \cap \overline{\partial\mathcal{K}^-}$ shared by two adjacent elements $\mathcal{K}^\pm$, with outward unit normal vectors $\mathbf{n}^\pm$, respectively, we define

$$\begin{aligned} [\![\tau]\!] &:= \tau^+ \cdot \mathbf{n}^+ + \tau^- \cdot \mathbf{n}^-, \qquad & [\![v]\!] &:= v^+\mathbf{n}^+ + v^-\mathbf{n}^-, \\ \{\!\{\tau\}\!\} &:= (\tau^+ + \tau^-)/2, & \{\!\{v\}\!\} &:= (v^+ + v^-)/2, \end{aligned}$$

where $\tau^\pm$ and $v^\pm$ denote the traces on $\partial\mathcal{K}^\pm$ taken from the interior of $\mathcal{K}^\pm$ of the (sufficiently regular) functions τ and v, respectively (cf. [5]). On a boundary face $\overline{F} = \overline{\partial\mathcal{K}} \cap \overline{\partial\Omega}$, we set $[\![\tau]\!] := \tau \cdot \mathbf{n}$, $[\![v]\!] := v\mathbf{n}$, $\{\!\{\tau\}\!\} := \tau$, and $\{\!\{v\}\!\} := v$.

We collect all interior (respectively, boundary) faces in the set $\mathcal{F}_h^I$ (respectively, $\mathcal{F}_h^B$), define $\mathcal{F}_h := \mathcal{F}_h^I \cup \mathcal{F}_h^B$, and introduce on $V_{h,p} \times V_{h,p}$ the following the bilinear form

$$\begin{aligned} \mathcal{A}(u,v) := &\sum_{\mathcal{K}\in\mathcal{T}_h} \int_\mathcal{K} \nabla u \cdot \nabla v \,\mathrm{dx} + \sum_{\mathcal{K}\in\mathcal{T}_h} \int_\mathcal{K} \nabla u \cdot \mathcal{R}([\![v]\!])\,\mathrm{dx} \\ &+ \sum_{\mathcal{K}\in\mathcal{T}_h} \int_\mathcal{K} \mathcal{R}([\![u]\!]) \cdot \nabla v \,\mathrm{dx} + \sum_{F\in\mathcal{F}_h} \int_F \alpha \frac{p^2}{|F|} [\![u]\!] \cdot [\![v]\!]\ \mathrm{ds}, \end{aligned}$$

where $\alpha > 0$ is a parameter at our disposal. The lifting operator $\mathcal{R}(\cdot)$ is defined as: $\mathcal{R}(\tau) := \sum_{F\in\mathcal{F}_h} r_F(\tau)$, where $r_F : [L^2(F)]^d \to [V_{h,p}]^d$ is given by

$$\int_\Omega r_F(\tau) \cdot \eta \,\mathrm{dx} := -\int_F \tau \cdot \{\!\{\eta\}\!\}\ \mathrm{ds} \quad \forall \eta \in [V_{h,p}]^d \quad \forall F \in \mathcal{F}_h.$$

The DG discretization of problem (1) reads:

$$\text{Find } u \in V_{h,p} \text{ such that} \qquad \mathcal{A}(u,v) = \int_\Omega fv \,\mathrm{dx} \quad \forall v \in V_{h,p}. \tag{2}$$

Let φ_j, $j = 1, \ldots, N_h^p := \dim(V_{h,p})$, be a set of basis functions that span $V_{h,p}$, then (2) can be written in the following equivalent form: Find $\mathbf{u} \in \mathbb{R}^{N_h^p}$ such that $\mathbf{Au} = \mathbf{f}$, where here (and in the following) we use the bold notation to denote the spaces of

degrees of freedom (vectors) and discrete linear operators (matrices). The following result provides an estimate for the spectral condition number of $\mathbf{A}$; we refer to [3] for the proof.

Proposition 1 ([3]). *For a set of basis functions which are orthonormal on the reference element $\widehat{\mathscr{K}} \subset \mathbb{R}^d$, $d = 2,3$, the condition number $\kappa(\boldsymbol{A})$ of the stiffness matrix $\boldsymbol{A}$ can be bounded by*

$$\kappa(\boldsymbol{A}) \lesssim \alpha \frac{p^4}{h^2}.$$

Remark 1. We are working, for the sake of simplicity, with the SIPG formulation proposed in [4], but the results shown in Proposition 1 and in Theorem 1 below also hold for a wide class of DG methods; we refer to [3] for details.

3 Two Level Non-overlapping Schwarz Preconditioners

In this section we introduce the non-overlapping Schwarz preconditioners.

Subdomain partition. We decompose the domain Ω into N non-overlapping subdomains Ω_i, i.e., $\overline{\Omega} = \cup_{i=1}^N \overline{\Omega}_i$. Next, we consider two levels of *nested* partitions of the domain Ω: (i) a coarse partition $\mathscr{T}_H$ (with mesh-size H); (ii) a fine partition $\mathscr{T}_h$ (with mesh-size h). We will suppose that the subdomain partition does not cut any element of $\mathscr{T}_H$ (and therefore of $\mathscr{T}_h$).

Local solvers. For $i = 1,\ldots,N$, we define the local DG spaces as

$$V_{h,p}^i := \{v \in L^2(\Omega_i) \ : \ v|_{\mathscr{K}} \circ F_{\mathscr{K}} \in \mathscr{M}^p(\widehat{\mathscr{K}}) \quad \forall\, \mathscr{K} \in \mathscr{T}_h, \mathscr{K} \subset \Omega_i\}.$$

Denoting by $R_i^{\mathrm{T}} : V_{h,p}^i \longrightarrow V_{h,p}$ the classical injection operator from $V_{h,p}^i$ to $V_{h,p}$, the local solvers $\mathscr{A}_i : V_{h,p}^i \times V_{h,p}^i \longrightarrow \mathbb{R}$ are defined as

$$\mathscr{A}_i(u_i, v_i) := \mathscr{A}(R_i^{\mathrm{T}} u_i, R_i^{\mathrm{T}} v_i) \quad \forall\, u_i, v_i \in V_{h,p}^i, \quad i = 1,\ldots,N. \tag{3}$$

Coarse solver. For an integer $0 \le q \le p$, we define the coarse space $V_{H,q}^0$ as

$$V_{H,q}^0 := \{v \in L^2(\Omega) \ : \ v|_{\mathscr{D}} \circ F_{\mathscr{D}} \in \mathbb{M}^{q_{\mathscr{D}}}(\widehat{\mathscr{K}}) \quad \forall\, \mathscr{D} \in \mathscr{T}_H\},$$

and the *coarse solver* $\mathscr{A}_0 : V_{H,q}^0 \times V_{H,q}^0 \longrightarrow \mathbb{R}$ as

$$\mathscr{A}_0(u_0, v_0) := \mathscr{A}(R_0^{\mathrm{T}} u_0, R_0^{\mathrm{T}} v_0) \quad \forall\, u_0, v_0 \in V_{H,q}^0, \tag{4}$$

where $R_0^{\mathrm{T}} : V_{H,q}^0 \longrightarrow V_{h,p}$ is the classical injection operator from $V_{H,q}^0$ to $V_{h,p}$.

Let the *local* projection operators be defined as

$$\begin{aligned}
&\widetilde{P}_i : V_{h,p} \to V_{h,p}^i : \quad \mathscr{A}_i(\widetilde{P}_i u, R_i^T v_i) := \mathscr{A}(u, R_i^T v_i) \quad \forall v_i \in V_{h,p}^i, \quad i = 1, \ldots, N, \\
&\widetilde{P}_0 : V_{h,p} \to V_{H,q}^0 : \quad \mathscr{A}_0(\widetilde{P}_0 u, R_0^T v_0) := \mathscr{A}(u, R_0^T v_0) \quad \forall v_0 \in V_{H,q}^0,
\end{aligned} \tag{5}$$

and define the projection operators as $P_i := R_i^T \widetilde{P}_i : V_{h,p} \longrightarrow V_{h,p}$, $i = 0, 1, \ldots, N$. The multiplicative Schwarz operator and its symmetrized variant are then defined as

$$P_{\text{mu}} := I - (I - P_N)(I - P_{N-1}) \cdots (I - P_0), \tag{6}$$

$$P_{\text{mu}}^{\text{S}} := I - (I - P_0)^T \cdots (I - P_N)^T (I - P_N) \cdots (I - P_0), \tag{7}$$

respectively (cf. [10]). The Schwarz method consists in solving either $P_{\text{mu}} u = g_{\text{mu}}$ or $P_{\text{mu}}^{\text{S}} u = g_{\text{mu}}^{\text{S}}$, for suitable right hand sides g_{mu} and g_{mu}^{S}, respectively. It can be shown that the operator defined in (7) is symmetric and positive definite; we therefore consider the conjugate gradient (CG) algorithm for the solution of $P_{\text{mu}}^{\text{S}} u = g_{\text{mu}}^{\text{S}}$. An estimate of the condition number of P_{mu}^{S} is

$$\kappa(P_{\text{mu}}^{\text{S}}) := \frac{\lambda_{\max}(P_{\text{mu}}^{\text{S}})}{\lambda_{\min}(P_{\text{mu}}^{\text{S}})},$$

where $\lambda_{\max}(P_{\text{mu}}^{\text{S}})$ and $\lambda_{\min}(P_{\text{mu}}^{\text{S}})$ are the extremal eigenvalues of the operator P_{mu}^{S}. On the other hand, the multiplicative operator P_{mu} is non-symmetric; we therefore consider a Richardson iteration applied to $P_{\text{mu}} u = g_{\text{mu}}$, and show that the norm of the error propagation operator $E_{\text{mu}} := (I - P_N)(I - P_{N-1}) \cdots (I - P_0)$ is strictly less than one, i.e.,

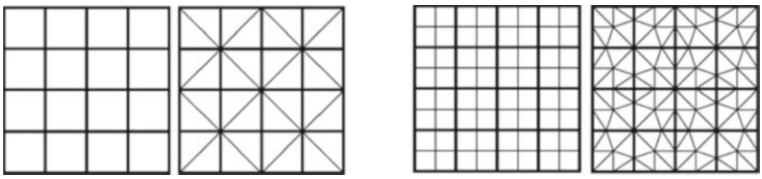
$$\|E_{\text{mu}}\|_{\mathscr{A}}^2 := \sup_{\substack{v \in V_{h,p} \\ v \neq 0}} \frac{\mathscr{A}(E_{\text{mu}} v, E_{\text{mu}} v)}{\mathscr{A}(v, v)} < 1,$$

and therefore a Richardson iteration applied to the preconditioned system converges. The following result provides a bound for the norm of the error propagation operator of the multiplicative Schwarz operator, and for the condition number of the symmetrized Schwarz operator (we refer to [3] for the proof).

Theorem 1 ([3]). *There exists constants $C_1, C_2 \geq 1$, independent of the mesh-size and the polynomial degree, such that*

$$\|E_{mu}\|_{\mathscr{A}}^2 \leq 1 - \frac{h}{C_1 \alpha p^2 H}, \quad \kappa(P_{mu}^S) \leq C_2 \alpha p^2 \frac{H}{h}.$$

Theorem 1 also guarantees that the multiplicative Schwarz method can be accelerated with the GMRES iterative solver. Indeed, according to [8], the GMRES method applied to the preconditioned system $P_{\text{mu}} u = g_{\text{mu}}$ does not stagnate (i.e., the iterative method makes some progress in reducing the residual at each iteration step) provided that: (i) $\|P_{mu}\|_{\mathscr{A}}$ is bounded; (ii) the symmetric part of P_{mu} is positive definite, i.e., there exists $c_p > 0$ such that $\mathscr{A}(v, P_{\text{mu}} v) >= c_p \mathscr{A}(v, v)$ for all $v \in V_{h,p}$. Condition (i) follows directly from the definition of P_{mu} and Theorem 1: $\|P_{mu}\|_{\mathscr{A}} = \|I - E_{\text{mu}}\|_{\mathscr{A}} \leq 1 + \|E_{\text{mu}}\|_{\mathscr{A}} < 2$. To prove condition (ii), it can be shown that



(a) (b)

Fig. 1. Initial Cartesian and triangular coarse and fine grids on a 16 subdomain partition. (**a**) Initial coarse grids (mesh-size H_0) and (**b**) initial fine grids (mesh-size h_0)

$$\mathscr{A}(P_{\rm mu}v,v) = \mathscr{A}(v,v) - \mathscr{A}(E_{\rm mu}v,v) \geq (1 - \|E_{\rm mu}\|_{\mathscr{A}})\,\mathscr{A}(v,v).$$

Therefore, condition (ii) holds true with $c_p = 1 - \|E_{\rm mu}\|_{\mathscr{A}}$ which is positive due to Theorem 1.

4 Numerical Results

In this section we present some numerical experiments to highlight the practical performance of the multiplicative and symmetrized non-overlapping Schwarz preconditioners. From the algebraic point of view, the Schwarz operators (6) and (7) can be written as the product of a suitable preconditioner, namely $\mathbf{B}_{\rm mu}$, $\mathbf{B}^{\rm S}_{\rm mu}$, respectively, and $\mathbf{A}$. Indeed, the local components can be constructed as $\mathbf{A}_i = \mathbf{R}_i\mathbf{A}\mathbf{R}_i^T$, see (3) for $i = 1\ldots,N$, and (4) for $i = 0$. From the definition (5) of the local projection $\widetilde{\mathbf{P}}_i = \mathbf{A}_i^{-1}\mathbf{R}_i\mathbf{A}$, and therefore $\mathbf{P}_i = \mathbf{R}_i^T\widetilde{\mathbf{P}}_i = \mathbf{R}_i^T\mathbf{A}_i^{-1}\mathbf{R}_i\mathbf{A}$. In practice, only the action of the preconditioner on a vector is needed. Algorithm 2 shows how to compute the action of $\mathbf{B}_{\rm mu}$ on a vector $\mathbf{x} \in \mathbb{R}^{N_h^p}$. Throughout this section we have set the

Algorithm 2 $\mathbf{z} = \mathbf{B}_{\rm mu}\mathbf{x}$

$\mathbf{z} = \mathbf{R}_0^T\mathbf{A}_0^{-1}\mathbf{R}_0\mathbf{x}$
for $i = 1 \rightarrow N$ **do**
  $\mathbf{z} = \mathbf{z} + \mathbf{R}_i^T\mathbf{A}_i^{-1}\mathbf{R}_i(\mathbf{x} - \mathbf{A}\mathbf{z})$
end for

penalty parameter $\alpha := 10$ (see (2)). We consider a subdomain partition consisting of $N = 16$ squares, and consider the initial Cartesian and unstructured triangular partitions shown in Fig. 1, and denote by H_0 and h_0 the corresponding initial coarse and fine mesh-sizes, respectively. We consider n successive global uniform refinements of these initial grids so that the resulting mesh-sizes are $H_n = H_0/2^n$ and $h_n = h_0/2^n$, with $n = 0,1,2,3$, respectively. The (relative) tolerance is set equal to 10^{-9} (respectively, 10^{-6}) for the CG (respectively, GMRES) iterative solver. We first address the performance of the multiplicative Schwarz preconditioner by keeping the mesh fixed, and varying the polynomial approximation degree p. In Table 1 we compare the GMRES iteration counts for both the preconditioned and non-preconditioned (in

Table 1. GMRES iteration counts. Multiplicative Schwarz preconditioner with a piecewise constant coarse solver ($q = 0$). Unstructured triangular grids.

	$h = h_0/2$ $H = H_0$	$h = h_0/4$ $H = H_0$	$h = h_0/4$ $H = H_0/2$
$p = 1$	23 (94)	33 (199)	25 (199)
$p = 2$	45 (259)	64 (540)	49 (540)
$p = 3$	66 (470)	93 (996)	74 (996)
$p = 4$	85 (713)	124 (1546)	97 (1546)
$p = 5$	105 (1004)	153 (2187)	123 (2187)
$p = 6$	124 (1342)	183 (2924)	144 (2924)
$p = 7$	143 (1727)	209 (3742)	167 (3742)
$p = 8$	162 (2148)	235 (4673)	189 (4673)
$p - rate$	0.93 (1.63)	0.88 (1.66)	0.93 (1.66)

parenthesis) systems, for different polynomial approximation degrees and different mesh configurations. These results have been obtained on unstructured triangular grids (cf. Fig. 1). Comparing the iteration counts of the preconditioned systems with the unpreconditioned ones for a fixed p, it is clear that the proposed preconditioner is very efficient. Indeed, we observe a reduction in the number of iterations needed to achieve convergence of around one order of magnitude when the proposed preconditioner is employed. The last row of Table 1 shows the computed growth rate in the number of iterations: we observe that the number of iterations needed to obtain convergence increases linearly as a function of p for the preconditioned system of equations, whereas this quantity grows almost quadratically for the non-preconditioned problem. In Fig. 2 we report the condition number estimates of the symmetrized Schwarz operator and the corresponding iteration counts versus the polynomial degree p. The solid lines refer to the mesh configuration $h = h_0/2$, $H = H_0$, whereas the dashed lines refer to the mesh configuration $h = h_0/4$, $H = H_0/2$. This set of numerical experiments has been obtained on Cartesian meshes, employing a piecewise linear coarse solver. As predicted by the theoretical estimates, the condition number of the preconditioned system grows quadratically as a function of p. Moreover, we clearly observe that, for fixed p, by refining both the fine and the coarse grid, but keeping the ratio of the fine and coarse mesh-sizes constant, the condition number (and therefore the number of iterations needed to obtain convergence) remains constant.

Next, we consider the performance of the symmetrized Schwarz preconditioner when varying the coarse and fine mesh-size, and keeping the polynomial approximation degree p fixed. In Table 2 (left) we report the condition number estimates for the symmetrized Schwarz operator employing piecewise biquadratic elements ($p = 2$) and a piecewise constant coarse solver ($q = 0$); whereas, in Table 2 (right) the analogous results obtained with piecewise bicubic elements ($p = 3$) and a piecewise linear coarse solver ($q = 1$) are shown. We clearly observe that the condition number grows

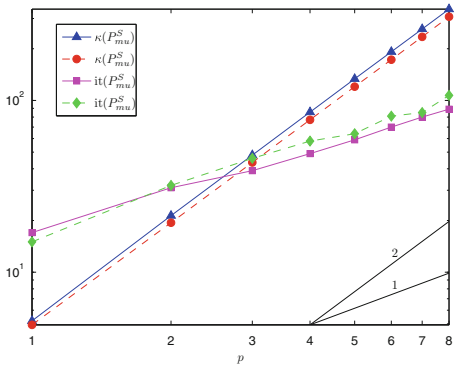


Fig. 2. Condition number estimates of the symmetrized Schwarz operator and corresponding iteration counts versus the polynomial degree p on Cartesian grids for different discretization steps (*solid line*: $h = h_0/2$, $H = H_0$; *dashed line* $h = h_0/4$, $H = H_0/2$). Piecewise linear coarse solver

Table 2. Condition number estimates for the symmetrized Schwarz operator with $p = 2$, $q = 0$ (left) and $p = 3$, $q = 1$ (right). Cartesian grids.

$h \downarrow H \rightarrow$	H_0	$H_0/2$	$H_0/4$	$H_0/8$	H_0	$H_0/2$	$H_0/4$	$H_0/8$
h_0	5.32e2	1.12e3	4.01e3	7.08e3	4.81e1	9.5925e1	1.92e2	3.91e2
$h_0/2$	2.74e2	4.71e2	2.80e3	5.59e3	2.14e1	4.35e1	8.70e1	1.75e2
$h_0/4$	–	2.60e2	1.18e3	3.42e3	–	2.09e1	4.24e1	8.44e1
$h_0/8$	–	–	3.45e2	1.75e3	–	–	2.05e1	4.26e1
$\kappa(\mathbf{A})$	2.88e5	1.18e6	4.89e6	1.99e7	7.44e5	2.81e6	1.11e7	4.55e7

as $O(Hh^{-1})$, as predicted by Theorem 1. Moreover, we clearly observe that employing a piecewise linear coarse solver ($q = 1$) rather than a piecewise constant coarse solver ($q = 0$) significantly improves the performance of the preconditioner. Indeed, comparing the condition number estimates of the preconditioned system with the analogous ones obtained for the non-preconditioned problem (last row of Table 2) we clearly observe that the condition number of the non-preconditioned system is reduced with respect to the condition number of the preconditioned system by approximately 5 orders of magnitude for $q = 1$ and 4 orders of magnitude for $q = 0$.

Bibliography

[1] P. F. Antonietti and B. Ayuso. Schwarz domain decomposition preconditioners for discontinuous Galerkin approximations of elliptic problems: non-overlapping case. *M2AN Math. Model. Numer. Anal.*, 41(1):21–54, 2007.

[2] P. F. Antonietti and B. Ayuso. Multiplicative Schwarz methods for discontinuous Galerkin approximations of elliptic problems. *M2AN Math. Model. Numer. Anal.*, 42(3):443–469, 2008.

[3] P. F. Antonietti and P. Houston. A class of domain decomposition preconditioners for *hp*-discontinuous Galerkin finite element methods. *J. Sci. Comp.*, 46(1):124–149, 2011.

[4] D. N. Arnold. An interior penalty finite element method with discontinuous elements. *SIAM J. Numer. Anal.*, 19(4):742–760, 1982.

[5] D. N. Arnold, F. Brezzi, B. Cockburn, and L. D. Marini. Unified analysis of discontinuous Galerkin methods for elliptic problems. *SIAM J. Numer. Anal.*, 39(5):1749–1779 (electronic), 2001/02.

[6] A. T. Barker, S. C. Brenner, P. Eun-Hee, and L.-Y. Sung. Two-level additive Schwarz preconditioners for a weakly over-penalized symmetric interior penalty method. *J. Sci. Comp.*, 47:27–49, 2011.

[7] S. C. Brenner and K. Wang. Two-level additive Schwarz preconditioners for C^0 interior penalty methods. *Numer. Math.*, 102(2):231–255, 2005.

[8] S. C. Eisenstat, H. C. Elman, and M. H. Schultz. Variational iterative methods for nonsymmetric systems of linear equations. *SIAM J. Numer. Anal.*, 20(2):345–357, 1983.

[9] X. Feng and O. A. Karakashian. Two-level additive Schwarz methods for a discontinuous Galerkin approximation of second order elliptic problems. *SIAM J. Numer. Anal.*, 39(4):1343–1365 (electronic), 2001.

[10] A. Toselli and O. Widlund. *Domain decomposition methods—algorithms and theory*, volume 34 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin, 2005.

A Block Solver for the Exponentially Fitted IIPG-0 Method

Blanca Ayuso de Dios[1], Ariel Lombardi[2], Paola Pietra[3], and Ludmil Zikatanov[4]

[1] Centre de Recerca Matemàtica, Barcelona, Spain. bayuso@crm.cat
[2] Departamento de Matemática, Universidad de Buenos Aires & CONICET, Buenos Aires, Argentina. aldoc7@dm.uba.ar
[3] IMATI-CNR, Pavia , Italy, pietra@imati.cnr.it
[4] Department of Mathematics, Penn State University, University Park, USA ludmil@psu.edu

Summary. We consider an exponentially fitted discontinuous Galerkin method for advection dominated problems and propose a block solver for the resulting linear systems. In the case of strong advection the solver is robust with respect to the advection direction and the number of unknowns.

1 Introduction

Let $\Omega \subset \mathbb{R}^2$ be a polygon, $f \in L^2(\Omega), g \in H^{1/2}(\partial\Omega)$ and let $\varepsilon > 0$ be constant. We consider the advection-diffusion problem

$$-\operatorname{div}(\varepsilon\nabla u - \beta u) = f \quad \text{in } \Omega, \qquad u = g \quad \text{on } \partial\Omega, \tag{1}$$

where $\beta \in [W^{1,\infty}(\Omega)]^2$ derives from a potential $\beta = \nabla\psi$. In applications to semiconductor devices, u is the electron density, ψ the electrostatic potential and the electric field $|\nabla\psi|$ might be fairly large in some parts of Ω, so that (1) becomes advection dominated. Its robust numerical approximation and the design of efficient solvers, are still a challenge. Exponential fitting [2] and discontinuous Galerkin (DG) are two approaches that have been combined in [3] to develop exponentially fitted DG methods (in primal and mixed formulation). In this note, we consider a variant of these schemes, based on the use of the Incomplete Interior Penalty IIPG-0 method and propose an efficient solver for the resulting linear systems.

The change of variable $\rho := e^{-\frac{\psi}{\varepsilon}} u$ in the problem (1) leads to

$$-\nabla\cdot(\kappa\nabla\rho) = f \text{ in } \Omega, \quad \rho = \chi \text{ on } \partial\Omega\,, \tag{2}$$

where $\kappa := \varepsilon e^{\frac{\psi}{\varepsilon}}$ and $\chi := e^{-\frac{\psi}{\varepsilon}} g$. An IIPG-0 approximation to (2) gives rise to the EF-IIPG-0 scheme for (1). We propose a block solver that uses ideas from [1] and reduce the solution to that of an exponentially fitted Crouziex-Raviart (CR) discretization,

R. Bank et al. (eds.), *Domain Decomposition Methods in Science and Engineering XX*, Lecture Notes in Computational Science and Engineering 91,
DOI 10.1007/978-3-642-35275-1_27,

which has much less degrees of freedom. The associated (CR) matrix is further reduced to an approximate block lower triangular form, which is efficiently solved by a block Gauss-Siedel algorithm.

In our description we focus on the case $\beta = \nabla\psi$ piecewise constant; although we include some numerical results for a more general case (cf. Test 2). Extensions of the method (allowing ψ to be discontinuous) and further analysis of the proposed solvers are topics of current research.

2 The Exponentially Fitted IIPG-0 Method

Let $\mathscr{T}_h$ be a shape-regular family of partitions of Ω into triangles T and let $h = \max_{T\in\mathscr{T}_h} h_T$ with h_T denoting the diameter of T for each $T \in \mathscr{T}_h$. We assume $\mathscr{T}_h$ does not contain hanging nodes. We denote by $\mathscr{E}_h^o$ and $\mathscr{E}_h^\partial$ the sets of all interior and boundary edges, respectively, and we set $\mathscr{E}_h = \mathscr{E}_h^o \cup \mathscr{E}_h^\partial$.

Let T^+ and T^- be two neighboring elements, and $\mathbf{n}^+$, $\mathbf{n}^-$ be their outward normal unit vectors, respectively ($\mathbf{n}^\pm = \mathbf{n}_{T^\pm}$). Let $\zeta^\pm$ and $\boldsymbol{\tau}^\pm$ be the restriction of ζ and $\boldsymbol{\tau}$ to $T^\pm$. We define the average and jump trace operators:

$$
\begin{aligned}
2\{\zeta\} &= (\zeta^+ + \zeta^-), \quad [\![\,\zeta\,]\!] = \zeta^+\mathbf{n}^+ + \zeta^-\mathbf{n}^- \qquad \text{on } E \in \mathscr{E}_h^o,\\
2\{\boldsymbol{\tau}\} &= (\boldsymbol{\tau}^+ + \boldsymbol{\tau}^-), \quad [\![\,\boldsymbol{\tau}\,]\!] = \boldsymbol{\tau}^+\cdot\mathbf{n}^+ + \boldsymbol{\tau}^-\cdot\mathbf{n}^- \qquad \text{on } E \in \mathscr{E}_h^o,
\end{aligned}
$$

and on $e \in \mathscr{E}_h^\partial$ we set $[\![\,\zeta\,]\!] = \zeta\mathbf{n}$ and $\{\boldsymbol{\tau}\} = \boldsymbol{\tau}$. We will also use the notation

$$
(u,w)_{\mathscr{T}_h} = \sum_{T\in\mathscr{T}_h}\int_T uwdx \qquad \langle u,w\rangle_{\mathscr{E}_h} = \sum_{e\in\mathscr{E}_h}\int_e uwds \quad \forall u,w,\in V^{DG}\,,
$$

where V^{DG} is the discontinuous linear finite element space defined by:

$$
V^{DG} = \left\{u \in L^2(\Omega)\ :\ u_{|_T}\ \in\ \mathbb{P}^1(T)\ \forall T \in \mathscr{T}_h\right\},
$$

Here, $\mathbb{P}^1(T)$ is the space of linear polynomials on T. Similarly, $\mathbb{P}^0(T)$ and $\mathbb{P}^0(e)$ are the spaces of constant polynomials on T and e, respectively. For each $e \in \mathscr{E}_h$, let $\mathscr{P}_e^0 : L^2(e) \mapsto \mathbb{P}^0(e)$ (resp. $\mathscr{P}_T^0 : L^2(T) \mapsto \mathbb{P}^0(T)$, for each $T \in \mathscr{T}_h$) be the L^2-orthogonal projections defined by

$$
\mathscr{P}_e^0(u) := \frac{1}{|e|}\int_e u, \quad \forall u \in L^2(e)\,, \quad \mathscr{P}_T^0(v) := \frac{1}{|T|}\int_T v, \quad \forall v \in L^2(T)\,.
$$

We denote by V^{CR} the classical Crouziex-Raviart (CR) space:

$$
V^{CR} = \left\{v \in L^2(\Omega)\ :\ v_{|_T} \in \mathbb{P}^1(T)\,\forall T \in \mathscr{T}_h \text{ and } \mathscr{P}_e^0[\![\,v\,]\!] = 0\ \forall e \in \mathscr{E}_h\right\}.
$$

Note that $v = 0$ at the midpoint m_e of each $e \in \mathscr{E}_h^\partial$. To represent the functions in V^{DG} we use the basis $\{\varphi_{e,T}\}_{T\in\mathscr{T}_h, e\in\mathscr{E}_h}$, defined by

$$\forall T \in \mathscr{T}_h \quad \varphi_{e,T}(x) \in \mathbb{P}^1(T) \quad e \subset \partial T \quad \varphi_{e,T}(m_{e'}) = \delta_{e,e'} \quad \forall e' \in \mathscr{E}_h \,. \tag{3}$$

In particular, any $w \in \mathbb{P}^1(T)$ can be written as $w = \sum_{e \subset \partial T} w(m_e)\varphi_{e,T}$.

We first consider the IIPG-0 approximation to the solution of (2): Find $\rho \in V^{DG}$ such that $\mathscr{A}(\rho, w) = (f, w)_{\mathscr{T}_h}$ for all $w \in V^{DG}$ with

$$\mathscr{A}(\rho, w) = (\kappa_T^* \nabla \rho, \nabla w)_{\mathscr{T}_h} - \langle \{\kappa_T^* \nabla \rho\}, [\![w]\!] \rangle_{\mathscr{E}_h} + \langle S_e \{[\![\rho]\!]\}, \mathscr{P}^0([\![w]\!]) \rangle_{\mathscr{E}_h} \,. \tag{4}$$

Here, S_e is the penalty parameter and $\kappa_T^* \in \mathbb{P}^0(T)$ the harmonic average approximation to $\kappa = \varepsilon e^{\psi/\varepsilon}$ both defined in [3] by:

$$\kappa_T^* := \frac{1}{\mathscr{P}_T^0(\kappa^{-1})} = \frac{\varepsilon}{\mathscr{P}_T^0(e^{-\frac{\psi}{\varepsilon}})} \,, \qquad S_e := \alpha_e h_e^{-1} \{\kappa_T^*\}_e \,, \tag{5}$$

Next, following [3] we introduce the local operator $\mathfrak{T} : V^{DG} \mapsto V^{DG}$ that approximates the change of variable introduced before (2):

$$\mathfrak{T} w := \sum_{T \in \mathscr{T}_h} (\mathfrak{T} w)|_T = \sum_{T \in \mathscr{T}_h} \sum_{e \subset \partial T} \mathscr{P}_e^0(e^{-\frac{\psi}{\varepsilon}}) w(m_e) \varphi_{e,T} \quad \forall w \in V^{DG} \,. \tag{6}$$

By setting $\rho := \mathfrak{T} u$ in (4), we finally get the EF-IIPG-0 approximation to (1):

Find $u_h \in V^{DG}$ s.t. $\mathscr{B}(u_h, w) := \mathscr{A}(\mathfrak{T} u_h, w) = (f, w)_{\mathscr{T}_h} \; \forall w \in V^{DG}$ with

$$\mathscr{B}(u, w) = (\kappa_T^* \nabla \mathfrak{T} u, \nabla w)_{\mathscr{T}_h} - \langle \{\kappa_T^* \nabla \mathfrak{T} u\}, [\![w]\!] \rangle_{\mathscr{E}_h} + \langle S_e \{[\![\mathfrak{T} u]\!]\}, \mathscr{P}^0 [\![w]\!] \rangle_{\mathscr{E}_h} \,. \tag{7}$$

It is important to emphasize that the use of harmonic average to approximate $\kappa = \varepsilon e^{\psi/\varepsilon}$ as defined in (5) together with the definition of the local approximation of the change of variables prevents possible overflows in the computations when $|\nabla \psi|$ is large and ε is small. (See [3] for further discussion).

Also, these two ingredients are essential to ensure that the resulting method has an automatic upwind mechanism built-in that allows for an accurate approximation of the solution of (1) in the advection dominated regime. We will discuss this in more detail in Sect. 3.

Prior to close this section, we define for each $e \in \mathscr{E}_h$ and $T \in \mathscr{T}_h$:

$$\psi_{m,e} := \min_{x \in e} \psi(x) \quad \psi_{m,T} := \min_{x \in T} \psi(x); \quad \psi_{m,T} \le \psi_{m,e} \text{ for } e \subset \partial T \,.$$

In the advection dominated regime $\varepsilon \ll |\beta| h = |\nabla \psi| h$

$$\mathscr{P}_T^0(e^{-(\psi/\varepsilon)}) \simeq \varepsilon^2 e^{-\frac{\psi_{m,T}}{\varepsilon}} \qquad \mathscr{P}_{e_i}^0(e^{-\psi/\varepsilon}) \simeq \varepsilon\, e^{-\frac{\psi_{m,e}}{\varepsilon}} \,. \tag{8}$$

The first of the above scalings together with the definitions in (5) imply

$$\kappa_T^* \simeq \frac{1}{\varepsilon} e^{\frac{\psi_{m,T}}{\varepsilon}} \,, \qquad S_e \simeq \frac{\alpha}{2\varepsilon} |e|^{-1} \exp\left(\frac{\psi_{m,T_1} + \psi_{m,T_2}}{\varepsilon}\right) \quad e = \partial T_1 \cap \partial T_2 \,. \tag{9}$$

3 Algebraic System and Properties

Let A and B be the operators associated to the bilinear forms $\mathscr{A}(\cdot,\cdot)$ (4) and $\mathscr{B}(\cdot,\cdot)$ (7), respectively. We denote by $\mathbb{A}$ and $\mathbb{B}$ their matrix representation in the basis $\{\varphi_{e,T}\}_{T\in\mathscr{T}_h, e\in\mathscr{E}_h}$ (3). In this basis, the operator $\mathfrak{T}$ defined in (6) is represented as a diagonal matrix, $\mathbb{D}$, and $\mathbb{B}=\mathbb{A}\mathbb{D}$. Thus, the approximation to (2) and (1) amounts to solve the linear systems (of dimension $2n_e-n_b$; with n_e and n_b being the cardinality of $\mathscr{E}_h$ and $\mathscr{E}_h^\partial$, respectively):

$$\mathbb{A}\boldsymbol{\rho}=\boldsymbol{F}\,,\quad \text{and}\quad \mathbb{D}\boldsymbol{u}=\boldsymbol{\rho}\qquad \text{or}\qquad \mathbb{B}\boldsymbol{u}=\widetilde{\boldsymbol{F}}\,, \tag{10}$$

where $\boldsymbol{\rho},\boldsymbol{u},\boldsymbol{F}$ and $\widetilde{\boldsymbol{F}}$ are the vector representations of ρ,u and the right hand sides of the approximate problems. From the definition (6) of $\mathfrak{T}$ it is easy to deduce the scaling of the entries of the diagonal matrix $\mathbb{D}=(d_{i,i})_{i=1}^{2n_e-n_b}$.

$$\mathbb{D}=(d_{i,j})_{i,j=1}^{2n_e-n_b}\quad d_{i,i}=\mathscr{P}^0_{e_i}(e^{-\psi/\varepsilon})\simeq\varepsilon\, e^{-\frac{\psi_{m,e}}{\varepsilon}}\,,\quad d_{i,j}\equiv 0\quad i\neq j\,.$$

We now revise a result from [1]:

Proposition 1. *Let $\mathscr{Z}\subset V^{DG}$ be the space defined by*

$$\mathscr{Z}=\left\{z\in L^2(\Omega)\,:\, z_{|_T}\in\mathbb{P}^1(T)\;\forall T\in\mathscr{T}_h \text{ and } \mathscr{P}^0_e\{v\}=0\;\forall e\in\mathscr{E}_h^o\right\}.$$

Then, for any $w\in V^{DG}$ there exists a unique $w^{cr}\in V^{CR}$ and a unique $w^z\in\mathscr{Z}$ such that $w=w^{cr}+w^z$, that is: $V^{DG}=V^{CR}\oplus\mathscr{Z}$. Moreover, $\mathscr{A}(w^{cr},w^z)=0\;\forall w^{cr}\in V^{CR}$, and $\forall\, w^z\in\mathscr{Z}$.

Proposition 1 provides a simple *change of basis* from $\{\varphi_{e,T}\}$ to canonical basis in V^{CR} and $\mathscr{Z}$ that results in the following algebraic structure for (10):

$$\boldsymbol{\rho}=\begin{bmatrix}\boldsymbol{\rho}^z\\ \boldsymbol{\rho}^{cr}\end{bmatrix},\qquad \mathbb{A}=\begin{bmatrix}\mathbb{A}^{zz} & \mathbf{0}\\ \mathbb{A}^{vz} & \mathbb{A}^{vv}\end{bmatrix},\qquad \mathbb{B}=\begin{bmatrix}\mathbb{B}^{zz} & 0\\ \mathbb{B}^{vz} & \mathbb{B}^{vv}\end{bmatrix}. \tag{11}$$

Due to the assumed continuity of ψ, $\mathbb{D}$ is still diagonal in this basis. The algebraic structure (11) suggests the following exact solver:

The solution $u=u^z+u^{cr}$ satisfying $\mathscr{B}(u,w)=(f,w)_{\mathscr{T}_h}$, for all $w\in V^{DG}$ is then obtained by

1. Solve for u^z: $\mathscr{B}(u^z,w^z)=(f,w^z)_{\mathscr{T}_h}\quad\forall w^z\in\mathscr{Z}$.
2. Solve for u^{cr}: $\mathscr{B}(u^{cr},w^{cr})=(f,w^{cr})_{\mathscr{T}_h}-\mathscr{B}(u^z,w^{cr})\quad\forall w^{cr}\in V^{CR}$.

Next, wet discuss how to solve efficiently each of the above steps.

Step 1: Solution in the $\mathscr{Z}$-space. In [1] it was shown that A^{zz} is a diagonal positive definite matrix. This is also true for $\mathbb{B}^{zz}$ since it is the product of two diagonal matrices. The continuity of ψ implies

$$\mathscr{B}(u^z, w^z) = \langle S_e \mathfrak{T}[\![u^z]\!], \mathscr{P}_e^0([\![w^z]\!])\rangle_{\mathscr{E}_h} \quad \forall\, u^z, w^z \in \mathscr{Z}\,. \tag{12}$$

Using (8) and (5) we observe that the entries of $\mathbb{B}^{zz}$ scale as:

$$\mathbb{B}^{zz} = (b_{i,j})_{i=1}^{n_e} \qquad b_{i,j} = S_{e_i}|e_i|d_j\delta_{i,j} \simeq \delta_{i,j}\frac{\alpha}{2}\, e^{-(\psi_{m,e}-\psi_{m,T_1}-\psi_{m,T_2})/\varepsilon}$$

which are always positive, so in particular $\mathbb{B}^{zz}$ it is also an M-matrix.

Step 2: Solution in V^{CR}. In [1] it was shown that the block $\mathbb{A}^{vv}$ coincides with the stiffness matrix of a CR discretization of (2), and so it is an s.p.d. matrix. However, this is no longer true for $\mathbb{B}^{vv}$ which is positive definite but non-symmetric.

$$\mathscr{B}(u^{cr}, w^{cr}) = (\kappa_T^* \nabla \mathfrak{T} u^{cr}, \nabla w^{cr})_{\mathscr{T}_h} \quad \forall\, u^{cr}, w^{cr} \in V^{CR}\,.$$

In principle, the sparsity pattern of $\mathbb{B}^{vv}$ is symmetric. Using (8) and (5), we find that the entries of the matrix scale as:

$$\mathbb{B}^{vv} = \left(b_{i,j}^{cr}\right)_{i,j}^{n_{cr}:=n_e-n_b} \qquad b_{i,j}^{cr} := \kappa_T^* \frac{|e_i||e_j|}{|T|}\mathbf{n}_{e_i}\cdot\mathbf{n}_{e_j} d_j \simeq e^{-\frac{(\psi_{m,e}-\psi_{m,T})}{\varepsilon}} \tag{13}$$

Since ψ is assumed to be piecewise linear, for each T, it attains its minimum (and also its maximum) at a vertex of T, say $\boldsymbol{x_0}$ and $\psi_{m,e}$ is attained at one of the vertex of the edge e, say $\boldsymbol{x_e}$. In particular, this implies that

$$\psi_{m,e} - \psi_{m,T} \approx \nabla\psi\cdot(\boldsymbol{x_e}-\boldsymbol{x_0}) = \beta\cdot(\boldsymbol{x_e}-\boldsymbol{x_0}) = \begin{cases} 0 & \boldsymbol{x_e}=\boldsymbol{x_0} \\ |\beta|h & \boldsymbol{x_e}\neq\boldsymbol{x_0} \end{cases}$$

Hence, in the advection dominated case $\varepsilon \ll |\beta|h$ some of the entries in (13) vanish (up to machine precision) for ε small; this is the automatic upwind mechanism intrinsic of the method. As a consequence, the sparsity pattern of $\mathbb{B}^{vv}$ is no longer symmetric and this can be exploited to re-order the unknowns so that $\mathbb{B}^{vv}$ can be reduced to block lower triangular form.

Notice also that for $\mathscr{T}_h$ acute, the block $\mathbb{A}^{vv}$ being the stiffness matrix of the Crouziex-Raviart approximation to (2), is an M-matrix. Hence, since the block $\mathbb{B}^{vv}$ is the product of a positive diagonal matrix and $\mathbb{A}^{vv}$, it will also be an M-matrix if the triangulation is acute (see [2]).

4 Block Gauss-Siedel Solver for V^{CR}-Block

We now consider re-orderings of the unknowns (dofs), which reduce $\mathbb{B}^{vv}$ to block lower triangular form. For such reduction, we use the algorithm from [4] which roughly amounts to *partitioning* the set of dofs into non-overlapping blocks. In the

strongly advection dominated case the size of the resulting blocks is small and a block Gauss-Seidel method is an efficient solver. Such techniques have been studied in [5] for conforming methods.

The idea is to consider the *directed* graph $\boldsymbol{G} = (\boldsymbol{V}, \boldsymbol{E})$ associated with $\mathbb{B}^{vv} \in \mathbb{R}^{n_{cr} \times n_{cr}}$; $\boldsymbol{G}$ has n_{cr} vertices labeled $\boldsymbol{V} = \{1, \ldots, n_{cr}\}$ and its set of *edges* $\boldsymbol{E}$ has cardinality equal to the number of nonzero entries[5] of $\mathbb{B}^{vv}$. By definition, $(i, j) \in \boldsymbol{E}$ *iff* $b^{cr}_{ij} \neq 0$. Note that in the advection dominated case, the built-in upwind mechanism results in a non-symmetric sparsity pattern for $\mathbb{B}^{vv}$ (see the last two paragraphs of Sect. 3). Thus, we may have $(i, j) \in \boldsymbol{E}$, while $(j, i) \notin \boldsymbol{E}$. Then, the problem of reducing $\mathbb{B}^{vv}$ to block lower triangular form of $\mathbb{B}^{vv}$ is equivalent to partitioning $\boldsymbol{G}$ as a union of strongly connected components.
Such partitioning induces non-overlapping partitioning of the set of dofs, $\boldsymbol{V} = \cup_{i=1}^{N_b} \omega_i$. For $i = 1, \ldots, N_b$, let m_i denote the cardinality of ω_i; let $\mathbb{I}_i \in \mathbb{R}^{n_{cr} \times m_i}$ be the matrix that is identity on dofs in ω_i and zero otherwise; and $\mathbb{B}^{vv}_i = \mathbb{I}_i^T \mathbb{B}^{vv} \mathbb{I}_i$ is the block corresponding to the dofs in ω_i. The block Gauss–Seidel algorithm reads: *Let* $\boldsymbol{u}^{cr}_0$ *be given, and assume* $\boldsymbol{u}^{cr}_k$ *has been obtained. Then* $\boldsymbol{u}^{cr}_{k+1}$ *is computed via:* For $i = 1, \ldots N_b$

$$\boldsymbol{u}^{cr}_{k+i/N_b} = \boldsymbol{u}^{cr}_{k+(i-1)/N_b} + \mathbb{I}_i (\mathbb{B}^{vv}_i)^{-1} \mathbb{I}_i^T \left(\boldsymbol{F} - \mathbb{B}^{vv} \boldsymbol{u}^{cr}_{k+(i-1)/N_b} \right) . \tag{14}$$

As we report in Sect. 5, the action of $(\mathbb{B}^{vv}_i)^{-1}$ can be computed exactly since in the advection dominated regime the size of the blocks $\mathbb{B}^{vv}_i$ is small.

5 Numerical Results

We present a set of numerical experiments to assess the performance of the proposed block solver. The tests refer to problem (2) with $\varepsilon = 10^{-3}, 10^{-5}, 10^{-7}$, and Ω is triangulated with a family of unstructured triangulations $\mathscr{T}_h$. In the tables given below $J = 1$ corresponds to the coarsest grid and each refined triangulation on level J, $J = 2, 3, 4$ is obtained by subdividing each of the $T \in \mathscr{T}_h$ on level $(J-1)$ into four congruent triangles. From the number of triangles n_T the total number of dofs for the DG approximation is $3n_T$.

Test 1. Boundary Layer: $\Omega = (-1, 1)^2$, $\beta = [1, 1]^t$, $n_T = 112$ for the coarsest mesh and f is such that the exact solution is given by

$$u(x, y) = \left(x + \frac{1 + e^{-2/\varepsilon} - 2e^{(x-1)/\varepsilon}}{1 - e^{-2/\varepsilon}} \right) \left(y + \frac{1 + e^{-2/\varepsilon} - 2e^{(y-1)/\varepsilon}}{1 - e^{-2/\varepsilon}} \right) .$$

Test 2. Rotating Flow: $\Omega = (-1, 1)^{\times}(0, 1)$, $f = 0$ and $\mathrm{curl}\beta \neq 0$,

[5] Each dof corresponds to a vertex in the graph; each nonzero entry to an edge.

$$\beta = \begin{bmatrix} 2y(1-x^2) \\ -2x(1-y^2) \end{bmatrix}^t \qquad g(x,y) = \begin{cases} 1+\tanh(10(2x+1)) & x \le 0,\ y = 0, \\ 0 & \text{elsewhere}\ . \end{cases}$$

We stress that this test does not fit in the simple description given here, and special care is required (see [3]). For the approximation, for each $T \in \mathscr{T}_h$, with barycenter (x_T, y_T), we use the approximation

$$\beta|_T \approx \nabla\psi|_T \qquad \text{with} \qquad \psi|_T = 2y_T(1-x_T^2)x - 2x_T(1-2y_T^2)y\ ,$$

and so ψ is discontinuous. The coarsest grid has $n_T = 224$ triangles. In Fig. 1 are

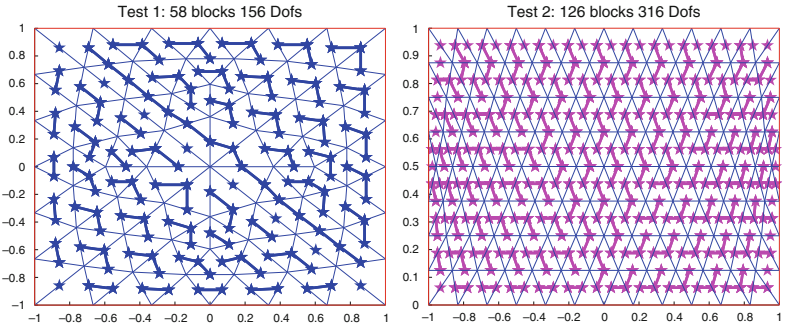


Fig. 1. Plot of the connected components (*blocks*) of $\mathbb{B}^{vv}$ created during Tarjan's algorithm: Test 1 with $\varepsilon = 10^{-5}$ (*left*); Test 2 with $\varepsilon = 10^{-7}$ (*right*)

represented the plot of the strongly connected components of the graph depicting the blocks for $\mathbb{B}^{vv}$ created during Tarjan's algorithm, on the coarsest meshes; for Test 1 with $\varepsilon = 10^{-5}$ (left figure) and for Test 2 with $\varepsilon = 10^{-7}$ (right figure). In Table 1 we report the number of blocks N_b created during Tarjan's algorithm; the maximum size of the largest such block (M_b); the average block size (n_{av}); and the number of block-Gauss-Seidel iterations. After Tarjan's algorithm is used to re-order the matrix $\mathbb{B}^{vv}$, we use the block Gauss-Seidel algorithm (14) where each small block is solved exactly. In the tests that we report here and also in all other similar tests that we have done (with similar advection dominance) the number of block-Gauss-Seidel iterations and the size of the blocks is uniformly bounded with respect to the number of dofs when the advection strongly dominates. Thus, the computational cost for one block Gauss-Seidel iteration in the advection dominated regime is the same as the cost of performing a fixed number of matrix vector multiplications and the algorithm is optimal in such regime.

Acknowledgments This work started while the first two authors were visiting the IMATI-CNR, Pavia in October 2010. Thanks go to the IMATI for the hospitality and support. First author was partially supported by MINECO grant MTM2011-27739-C04-04 and GENCAT grant 2009-SGR-345. Second author was supported by CONICET, Argentina. The fourth author was supported in part by National Science Foundation NSF-DMS 0810982.

Test 1

ε \ J		1	2	3	4
10^{-3}	N_b	44	150	484	1182
	M_b	23	47	95	191
	n_{av}	3.55	4.32	5.45	9.02
	iters	7	19	43	166
10^{-5}	N_b	50	210	866	3474
	M_b	23	47	95	191
	n_{av}	3.12	3.08	3.05	3.07
	iters	4	4	4	14
10^{-7}	N_b	50	210	866	3522
	M_b	23	47	95	191
	n_{av}	3.12	3.08	3.05	3.03
	iters	4	4	4	4

Test 2

ε \ J		1	2	3	4
10^{-3}	N_b	31	1	1	1
	M_b	211	1304	5296	21344
	n_{av}	10.19	1304	5296	21344
	iters	10	1	1	1
10^{-5}	N_b	122	468	1822	7106
	M_b	4	4	7	37
	n_{av}	2.59	2.78	2.91	3.00
	iters	4	4	7	24
10^{-7}	N_b	122	468	1832	7247
	M_b	4	4	4	6
	n_{av}	2.59	2.78	2.89	2.95
	iters	4	4	4	4

Table 1. Number of blocks (N_b) created during the Tarjan's ordering algorithm, size of largest block (M_b), average size of blocks (n_{av}) and number of block-Gauss-Seidel iterations (iters) for Test 1 (left) and Test 2 (right).

Bibliography

[1] Blanca Ayuso de Dios and Ludmil Zikatanov. Uniformly convergent iterative methods for discontinuous Galerkin discretizations. *J. Sci. Comput.*, 40(1–3): 4–36, 2009.

[2] F. Brezzi, L. D. Marini, S. Micheletti, P. Pietra, R. Sacco, and S. Wang. Discretization of semiconductor device problems. I. In *Handbook of numerical analysis. Vol. XIII*, pages 317–441. North-Holland, Amsterdam, 2005.

[3] Ariel Lombardi and P. Pietra. Exponentially fitted discontinuous galerkin schemes for singularly perturbed problems. numerical methods for partial differential equations. 2011. (to appear) doi: 10.1002/num.20701.

[4] Robert Tarjan. Depth-first search and linear graph algorithms. *SIAM J. Comput.*, 1(2):146–160, 1972.

[5] Feng Wang and Jinchao Xu. A crosswind block iterative method for convection-dominated problems. *SIAM J. Sci. Comput.*, 21(2):620–645, 1999.

A Nonoverlapping DD Preconditioner for a Weakly Over-Penalized Symmetric Interior Penalty Method

Andrew T. Barker[1], Susanne C. Brenner[2], Eun-Hee Park[3], and Li-Yeng Sung[4]

[1] Department of Mathematics and Center for Computation and Technology, Louisiana State University, Baton Rouge, LA 70803, USA andrewb@math.lsu.edu
[2] Department of Mathematics and Center for Computation and Technology, Louisiana State University, Baton Rouge, LA 70803, USA brenner@math.lsu.edu
[3] Department of Mathematics and Center for Computation and Technology, Louisiana State University, Baton Rouge, LA 70803, USA epark2@math.lsu.edu
[4] Department of Mathematics and Center for Computation and Technology, Louisiana State University, Baton Rouge, LA 70803, USA sung@math.lsu.edu

1 Introduction

In this paper we present a nonoverlapping domain decomposition preconditioner for a weakly over-penalized symmetric interior penalty method that is based on balancing domain decomposition by constraints (BDDC) methodology (cf. [2, 5, 7, 8]). The full analysis of the preconditioner can be found in [4].

Let Ω be a bounded polygonal domain in $\mathbb{R}^2$ and $f \in L_2(\Omega)$. Consider the following model problem:
Find $u \in H_0^1(\Omega)$ such that

$$\int_\Omega \nabla u \cdot \nabla v \, dx = \int_\Omega f v \, dx \qquad \forall v \in H_0^1(\Omega). \tag{1}$$

Let $\mathscr{T}_h$ be a quasi-uniform triangulation of Ω, where the mesh parameter h measures the maximum diameter of the triangles in $\mathscr{T}_h$, and let

$$V_h = \{v \in L_2(\Omega) : v|_T \in P_1(T) \quad \forall T \in \mathscr{T}_h\}$$

be the discontinuous P_1 finite element function space associated with $\mathscr{T}_h$. The model problem (1) can be discretized by the following weakly over-penalized symmetric interior penalty (WOPSIP) method (cf. [3, 9]):
Find $u_h \in V_h$ such that

$$a_h(u_h, v) = \int_\Omega f v \, dx \qquad v \in V_h,$$

where

R. Bank et al. (eds.), *Domain Decomposition Methods in Science and Engineering XX*, Lecture Notes in Computational Science and Engineering 91,
DOI 10.1007/978-3-642-35275-1_28, © Springer-Verlag Berlin Heidelberg 2013

$$a_h(v,w) = \sum_{T\in\mathscr{T}_h} \int_T \nabla v \cdot \nabla w \, dx + \sum_{e\in\mathscr{E}_h} \frac{1}{|e|^3} \int_e \Pi_e^0 [\![v]\!] \cdot \Pi_e^0 [\![w]\!] \, ds, \tag{2}$$

$\mathscr{E}_h$ is the set of the edges of $\mathscr{T}_h$, $|e|$ is the length of the edge e, $[\![v]\!]$ denotes the jump of v across the edges, and Π_e^0 is the orthogonal projection from $[L_2(e)]^2$ onto $[P_0(e)]^2$. $P_0(e)$ denotes the space of constant functions on the edge e.

For simplicity in presentation, we consider the Poisson model on conforming meshes. But the results can be extended to heterogeneous elliptic problems on nonconforming meshes (cf. [4]). We note that BDDC technique was used in [6] to couple conforming finite element spaces from different subdomains that allows nonmatching meshes across subdomain boundaries, where condition number estimates independent of the coefficients were obtained for heterogeneous elliptic problems. The main difference between [6] and this paper is that the finite element functions in this paper can be discontinuous at the element boundaries.

The rest of the paper is organized as follows. In Sect. 2 we introduce a subspace decomposition. We then design a BDDC preconditioner for the reduced problem in Sect. 3. The condition number estimate is also presented. In Sect. 4 we report numerical results that illustrate the performance of the proposed preconditioner and confirm the theoretical estimates.

Throughout the paper we will use $A \lesssim B$ and $A \gtrsim B$ to represent the statements that $A \leq (\text{constant})B$ and $A \geq (\text{constant})B$, where the positive constant is independent of the mesh size, the subdomain size, and the number of subdomains. The statement $A \approx B$ is equivalent to $A \lesssim B$ and $A \gtrsim B$.

2 A Subspace Decomposition

In this section we propose an intermediate preconditioner for the WOPSIP method, which is based on a subspace decomposition.

Let $\Omega_1, \ldots, \Omega_J$ be a nonoverlapping partition of Ω aligned with $\mathscr{T}_h$ and $\Gamma = \left(\bigcup_{j=1}^J \partial\Omega_j\right) \setminus \partial\Omega$ be the interface of the subdomains. We assume that the subdomains are shape regular polygons (cf. [1, Sect. 7.5]). We denote the diameter of Ω_j by H_j and define H to be $\max_{1\leq j\leq J} H_j$. $\mathscr{E}_{h,\Gamma}$ is the subset of $\mathscr{E}_h$ containing the edges on Γ.

First we decompose V_h into two subspaces as follows:

$$V_h = V_{h,C} \oplus V_{h,D},$$

where

$$\begin{aligned} V_{h,C} = \{ & v \in V_h : [\![v]\!] = 0 \text{ at the midpoints of the edges on the boundaries} \\ & \text{of the subdomains} \}, \\ V_{h,D} = \{ & v \in V_h : \{\!\{ v \}\!\} = 0 \text{ at the midpoints of the edges in } \mathscr{E}_{h,\Gamma} \text{ and} \\ & v = 0 \text{ at the midpoints of the edges in } \Omega \setminus \Gamma \}. \end{aligned}$$

Here $\{\!\{v\}\!\}$ denotes the average of v from the two sides of an edge in $\mathscr{E}_{h,\Gamma}$.

Let $A_h : V_h \longrightarrow V_h'$ be the symmetric positive-definite (SPD) operator defined by

$$\langle A_h v, w\rangle = a_h(v,w) \qquad \forall v, w \in V_h,$$

where $\langle\cdot,\cdot\rangle$ is the canonical bilinear form between a vector space and its dual. Similarly, we define $A_{h,D} : V_{h,D} \longrightarrow V_{h,D}'$ and $A_{h,C} : V_{h,C} \longrightarrow V_{h,C}'$ by

$$\langle A_{h,D} v, w\rangle = a_h(v,w) \qquad \forall v, w \in V_{h,D}, \tag{3}$$
$$\langle A_{h,C} v, w\rangle = a_h(v,w) \qquad \forall v, w \in V_{h,C}. \tag{4}$$

Given any $v \in V_h$, we have a unique decomposition $v = v_D + v_C$ where $v_D \in V_{h,D}$ and $v_C \in V_{h,C}$. Then based on the definitions of the subspaces $V_{h,D}$ and $V_{h,C}$, it can be shown that

$$\langle A_h v, v\rangle \approx \langle A_{h,D} v_D, v_D\rangle + \langle A_{h,C} v_C, v_C\rangle \qquad \forall v \in V_h. \tag{5}$$

Remark 1. Since functions in $V_{h,C}$ are continuous at the midpoints of the edges in $\mathscr{E}_{h,\Gamma}$, we have

$$a_h(v,w) = \sum_{j=1}^{J} a_{h,j}(v_j, w_j) \qquad \forall v, w \in V_{h,C}, \tag{6}$$

where $v_j = v\big|_{\Omega_j}$, $w_j = w\big|_{\Omega_j}$ and

$$a_{h,j}(v_j, w_j) = \sum_{\substack{T\in\mathscr{T}_h \\ T\subset\Omega_j}} \int_T \nabla v_j \cdot \nabla w_j \, dx + \sum_{\substack{e\in\mathscr{E}_h \\ e\subset\Omega_j}} \frac{1}{|e|^3} \int_e \Pi_e^0[\![v_j]\!] \cdot \Pi_e^0[\![w_j]\!] \, ds. \tag{7}$$

Note that the second sum on the right-hand side of (7) is over the edges interior to Ω_j and therefore $a_{h,j}(\cdot,\cdot)$ is a localized bilinear form. The introduction of the subspace decomposition where the bilinear form can be localized as shown in (6) and (7) is the key ingredient in designing our preconditioner in Sect. 3.

Next we decompose $V_{h,C}$ into two subspaces $V_{h,C}(\Omega\setminus\Gamma)$ and $V_{h,C}(\Gamma)$ defined as follows:

$$V_{h,C}(\Omega\setminus\Gamma) = \{v \in V_{h,C} : v = 0 \text{ at all the midpoints of the edges in } \mathscr{E}_{h,\Gamma}\},$$
$$V_{h,C}(\Gamma) = \{v \in V_{h,C} : a_h(v,w) = 0 \quad \forall w \in V_{h,C}(\Omega\setminus\Gamma)\}.$$

The space $V_{h,C}(\Gamma)$ is the space of discrete harmonic functions, which are uniquely determined by their values at the midpoints of the edges in $\mathscr{E}_{h,\Gamma}$.

Let the SPD operators $A_{h,\Omega\setminus\Gamma} : V_{h,C}(\Omega\setminus\Gamma) \longrightarrow V_{h,C}(\Omega\setminus\Gamma)'$ and $S_h : V_{h,C}(\Gamma) \longrightarrow V_{h,C}(\Gamma)'$ be defined by

$$\langle A_{h,\Omega\setminus\Gamma} v, w\rangle = a_h(v,w) \qquad \forall v, w \in V_{h,C}(\Omega\setminus\Gamma),$$
$$\langle S_h v, w\rangle = a_h(v,w) \qquad \forall v, w \in V_{h,C}(\Gamma).$$

Note that given any $v_C \in V_{h,C}$, we have a unique decomposition $v_C = v_{C,\Omega\setminus\Gamma} + v_{C,\Gamma}$ where $v_{C,\Omega\setminus\Gamma} \in V_{h,C}(\Omega\setminus\Gamma)$ and $v_{C,\Gamma} \in V_{h,C}(\Gamma)$. It follows from the definitions of $V_{h,C}(\Omega\setminus\Gamma)$ and $V_{h,C}(\Gamma)$ that

$$\langle A_{h,C}v_C, v_C\rangle = \langle A_{h,\Omega\setminus\Gamma}v_{C,\Omega\setminus\Gamma}, v_{C,\Omega\setminus\Gamma}\rangle + \langle S_h v_{C,\Gamma}, v_{C,\Gamma}\rangle \qquad \forall v_C \in V_{h,C}. \tag{8}$$

Based on the relations (5) and (8), we define a preconditioner $B_1 : V_h{}' \longrightarrow V_h$ for A_h by

$$B_1 = I_D A_{h,D}^{-1} I_D^t + I_{h,\Omega\setminus\Gamma} A_{h,\Omega\setminus\Gamma}^{-1} I_{h,\Omega\setminus\Gamma}^t + I_\Gamma S_h^{-1} I_\Gamma^t,$$

where $I_D : V_{h,D} \longrightarrow V_h$, $I_{h,\Omega\setminus\Gamma} : V_{h,C}(\Omega\setminus\Gamma) \longrightarrow V_h$, and $I_\Gamma : V_{h,C}(\Gamma) \longrightarrow V_h$ are natural injections.

It follows from (5) and (8) that

$$\kappa(B_1A_h) = \frac{\lambda_{\max}(B_1A_h)}{\lambda_{\min}(B_1A_h)} \approx 1. \tag{9}$$

Remark 2. Let us observe the properties of the preconditioner B_1 from the implementational point of view. First it is easy to implement the solve $A_{h,D}^{-1}$ because $A_{h,D}$ is a block diagonal matrix with small blocks. Next in view of (6) and (7), the solve $A_{h,\Omega\setminus\Gamma}^{-1}$ can be implemented by solving independent subdomain problems in parallel. On the other hand, noting that S_h is a global solve, we need to design a good preconditioner for S_h in order to obtain a good parallel preconditioner for A_h.

3 A BDDC Preconditioner

In this section we propose a preconditioner for the Schur complement operator S_h based on the BDDC methodology.

Let $V_{h,j}$ be the space of discontinuous P_1 finite element functions on Ω_j that vanish at the midpoints of the edges on $\partial\Omega_j \cap \partial\Omega$, and $V_h(\Omega_j)$ be the subspace of $V_{h,j}$ whose members vanish at the midpoints of the edges on $\partial\Omega_j$. We denote by $\mathscr{H}_j$ the space of local discrete harmonic functions defined by

$$\mathscr{H}_j = \left\{v \in V_{h,j} : a_{h,j}(v,w) = 0 \quad \forall w \in V_h(\Omega_j)\right\}.$$

The space $\mathscr{H}_m$ is defined by gluing the spaces $\mathscr{H}_j$ together along the interface Γ through enforcing the continuity of the mean values on the common edges of subdomains:

$$\begin{aligned}\mathscr{H}_m = \{v \in L_2(\Omega) : v_j = v|_{\Omega_j} \in \mathscr{H}_j \text{ for } 1 \le j \le J \\ \text{and } \int_{\partial\Omega_j\cap\partial\Omega_k} v_j\,ds = \int_{\partial\Omega_j\cap\partial\Omega_k} v_k\,ds \text{ for } 1 \le j,k \le J\},\end{aligned}$$

and we equip $\mathscr{H}_m$ with the bilinear form

$$a_h^m(v,w) = \sum_{1\le j\le J} a_{h,j}(v_j, w_j).$$

Let $\mathscr{E}_H$ be the set of the edges of the subdomains $\Omega_1, \cdots, \Omega_J$. The BDDC preconditioner is based on a decomposition of $\mathscr{H}_m$ into orthogonal subspaces with respect to $a_h^m(\cdot,\cdot)$:

$$\mathscr{H}_m = \mathring{\mathscr{H}} \oplus \mathscr{H}_0, \tag{10}$$

where

$$\mathring{\mathscr{H}} = \left\{ v \in \mathscr{H}_m : \int_E v\, ds = 0 \quad \forall E \in \mathscr{E}_H \right\}$$

and

$$\mathscr{H}_0 = \left\{ v \in \mathscr{H}_m : a_h^m(v,w) = 0 \quad \forall w \in \mathring{\mathscr{H}} \right\}. \tag{11}$$

Then we equip $\mathscr{H}_0$ and the localized subspaces $\mathring{\mathscr{H}}_j$ $(1 \le j \le J)$ of $\mathring{\mathscr{H}}$:

$$\mathring{\mathscr{H}}_j = \left\{ v \in \mathscr{H}_j : \int_E v\, ds = 0 \text{ for all the edges } E \text{ of } \Omega_j \right\},$$

with the SPD operators $S_0 : \mathscr{H}_0 \longrightarrow \mathscr{H}_0'$ and $S_j : \mathring{\mathscr{H}}_j \longrightarrow \mathring{\mathscr{H}}_j'$ defined by

$$\langle S_0 v, w \rangle = a_h^m(v,w) \qquad \forall v, w \in \mathscr{H}_0, \tag{12}$$

$$\langle S_j v, w \rangle = a_{h,j}(v,w) \qquad \forall v, w \in \mathring{\mathscr{H}}_j. \tag{13}$$

Note that $V_{h,C}(\Gamma)$ is a subspace of $\mathscr{H}_m$ and there exists a projection $P_\Gamma : \mathscr{H}_m \to V_{h,C}(\Gamma)$ defined by averaging:

$$(P_\Gamma v)(m_e) = \{\!\{ v \}\!\}(m_e) \quad \forall e \in \mathscr{E}_{h,\Gamma},$$

where m_e is the midpoint of e. The operator P_Γ connects the BDDC preconditioner based on $\mathscr{H}_m$ to the Schur complement operator S_h on $V_{h,C}(\Gamma)$.

We can now define the BDDC preconditioner $B_{BDDC} : V_{h,C}(\Gamma)' \longrightarrow V_{h,C}(\Gamma)$ for the Schur complement operator $S_h : V_{h,C}(\Gamma) \longrightarrow V_{h,C}(\Gamma)'$ as follows:

$$B_{BDDC} = (P_\Gamma I_0)\, S_0^{-1}\, (P_\Gamma I_0)^t + \sum_{j=1}^{J} (P_\Gamma \mathbb{E}_j)\, S_j^{-1}\, (P_\Gamma \mathbb{E}_j)^t,$$

where I_0 is the natural injection of $\mathscr{H}_0$ into $\mathscr{H}_m$ and $\mathbb{E}_j : \mathring{\mathscr{H}}_j \longrightarrow \mathring{\mathscr{H}}$ is the trivial extension defined by

$$\mathbb{E}_j \mathring{v}_j = \begin{cases} \mathring{v}_j & \text{on } \Omega_j \\ 0 & \text{on } \Omega \setminus \Omega_j \end{cases} \qquad \forall \mathring{v}_j \in \mathring{\mathscr{H}}_j.$$

We then obtain the preconditioner $B_2 : V_h' \longrightarrow V_h$ for A_h by replacing the global solve S_h^{-1} in (2) with the preconditioner B_{BDDC}:

$$B_2 = I_D A_{h,D}^{-1} I_D^t + I_{h,\Omega\setminus\Gamma} A_{h,\Omega\setminus\Gamma}^{-1} I_{h,\Omega\setminus\Gamma}^t + I_\Gamma B_{BDDC} I_\Gamma^t.$$

We can analyze the condition number of $B_{BDDC} S_h$ by the theory of additive Schwarz preconditioners (cf. [1, 10, 11], and the references therein). The proof of the following result can be found in [4].

Lemma 1. *We have the following bounds for the eigenvalues of* $B_{BDDC}S_h$

$$\lambda_{\min}(B_{BDDC}S_h) \geq 1,$$
$$\lambda_{\max}(B_{BDDC}S_h) \lesssim \left(1+\ln\frac{H}{h}\right)^2.$$

Combining (5), (8) and Lemma 1, we have the following estimate of the condition number of the preconditioned system B_2A_h.

Theorem 1. *There exists a positive constant C, independent of h, H and J, such that*

$$\kappa(B_2A_h) = \frac{\lambda_{\max}(B_2A_h)}{\lambda_{\min}(B_2A_h)} \leq C\left(1+\ln\frac{H}{h}\right)^2.$$

4 Numerical Results

In this section we present some numerical results that illustrate the performance of the preconditioners B_1 and B_2.

We consider the model problem (1) on the unit square $(0,1)^2$ with the exact solution $u(x,y) = y(1-y)\sin(\pi x)$. We use a uniform triangulation $\mathscr{T}_h$ of isosceles right triangles, where the mesh parameter h represents the length of the horizontal/vertical edges. The domain Ω is divided into J nonoverlapping squares aligned with $\mathscr{T}_h$ and the length of the horizontal/vertical edges of the squares is denoted by H. The discrete problem obtained by the WOPSIP method is solved by the preconditioned conjugate gradient method. The iteration is stopped when the relative residual is less than 10^{-6}.

Numerical results for the preconditioners B_1 and B_2 are presented in Table 1, which confirm the theoretical estimates in (9) and Theorem 1.

Table 1. Results for the preconditioners B_1 and B_2 with $J = 4^2$

h	H/h	B_1A_h			B_2A_h		
		κ	$\lambda_{\min}$	$\lambda_{\max}$	κ	$\lambda_{\min}$	$\lambda_{\max}$
2^{-3}	2	1.4206	8.2624e-1	1.1738	1.4478	8.2623e-1	1.1962
2^{-4}	4	1.1916	9.1258e-1	1.0874	1.7782	9.1300e-1	1.6235
2^{-5}	8	1.0919	9.5608e-1	1.0439	2.3215	9.5673e-1	2.2211
2^{-6}	16	1.0433	9.7880e-1	1.0212	3.0490	9.7994e-1	2.9879

We present in Table 2 the iteration counts and total time to solution for a parallel implementation of our preconditioner. For comparison, results on a single processor of the same machine without preconditioning are also presented for $J = 1$. The three operations $A_{h,D}^{-1}, A_{h,\Omega\setminus\Gamma}^{-1}$, and B_{BDDC} are performed one after the other, sequentially,

but each of these operators is evaluated in parallel on the decomposed domain with one subdomain per processor. Iteration counts are consistent with our theory and confirm again that the method is scalable, and the running times show good parallel speedup for large problems.

Table 2. Parallel performance of the preconditioner B_2

h	$J=1$		$J=4^2, H=2^{-2}$		$J=8^2, H=2^{-3}$		$J=16^2, H=2^{-4}$	
	Its	Wall clock time	Its	Wall clock time	Its	Wall clock time	Its	Wall clock time
2^{-6}	235	0.46	7	0.37	7	0.5	5	1.14
2^{-7}	450	3.75	8	2.22	8	1.06	6	1.96
2^{-8}	884	35.45	9	20.12	8	4.35	6	2.71
2^{-9}	1786	319.0	8	126.15	8	27.15	7	7.81

The numbers $\kappa(B_2A_h)/(1+\ln(H/h))^2$ and $\kappa(B_{BDDC}S_h)/(1+\ln(H/h))^2$ are plotted against H/h in Fig. 1. As H/h increases these two numbers settle down to around 0.2, which indicates that the estimates in Lemma 1 and Theorem 1 are sharp.

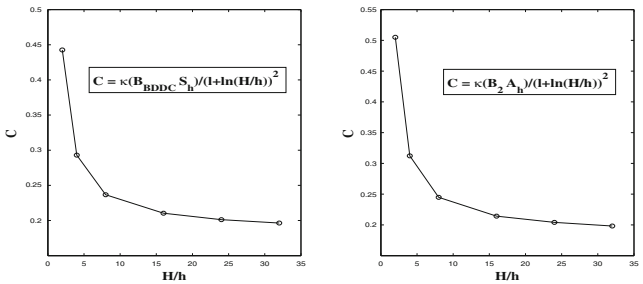


Fig. 1. *Left figure*: the behavior of $C=\kappa(B_{BDDC}S_h)/(1+\ln(H/h))^2$ for the BDDC preconditioner; *right figure*: the behavior of $C=\kappa(B_2A_h)/(1+\ln(H/h))^2$ for the preconditioner B_2

Acknowledgments The work of the first author was supported in part by the National Science Foundation under Grant No. DMS-07-39382. The work of the second and forth authors was supported in part by the National Science Foundation under Grant No. DMS-10-16332. The work of the third author was supported in part by the National Research Foundation of Korea under Grant No. NRF-2009-352-C00009. The work of all four authors was supported in part by the Institute for Mathematics and its Applications with funds provided by the National Science Foundation.

Bibliography

[1] S. C. Brenner and L. R. Scott. *The mathematical theory of finite element methods (Third Edition)*. Springer, 2008.
[2] S. C. Brenner and L.-Y. Sung. BDDC and FETI-DP without matrices or vectors. *Comput. Methods Appl. Mech. Engrg.*, 196:1429–1435, 2007.
[3] S. C. Brenner, L. Owens, and L.-Y. Sung. A weakly over-penalized symmetric interior penalty method. *Electron. Trans. Numer. Anal.*, 30:107–127, 2008.
[4] S. C. Brenner, E.-H. Park, and L.-Y. Sung. A BDDC preconditioner for a weakly over-penalized symmetric interior penalty method. *preprint*, 2011.
[5] C. R. Dohrmann. A preconditioner for substructuring based on constrained energy minimization. *SIAM J. Sci. Comput.*, 25:246–258, 2003.
[6] M. Dryja, J. Galvis, and M. Sarkis. BDDC methods for discontinuous Galerkin discretization of elliptic problems. *J. Complex.*, 23:715–739, 2007.
[7] J. Li and O. B. Widlund. FETI-DP, BDDC, and block Cholesky methods. *Internat. J. Numer. Methods Engrg.*, 66:250–271, 2006.
[8] J. Mandel and C. R. Dohrmann. Convergence of a balancing domain decomposition by constraints and energy minimization. *Numer. Linear Algebra Appl.*, 10:639–659, 2003.
[9] L. Owens. *Multigrid methods for weakly over-penalized interior penalty methods*. Ph.D. thesis, University of South Carolina, 2007.
[10] B. Smith, P. Bjørstad, and W. Gropp. *Domain Decomposition*. Cambridge University Press, Cambridge, 1996.
[11] A. Toselli and O. B. Widlund. *Domain Decomposition Methods - Algorithms and Theory*. Springer-Verlag, Berlin, 2005.

Sharp Condition Number Estimates for the Symmetric 2-Lagrange Multiplier Method

Stephen W. Drury[1] and Sébastien Loisel[2]

[1] Dept. of Mathematics, McGill University, 805 Sherbrooke Street West, Montreal, Quebec, Canada, H3A 2K6, drury@math.mcgill.ca
[2] Dept. of Mathematics, Heriot-Watt University, Edinburgh EH14 4AS, United Kingdom, S.Loisel@hw.ac.uk

Summary. Domain decomposition methods are used to find the numerical solution of large boundary value problems in parallel. In optimized domain decomposition methods, one solves a Robin subproblem on each subdomain, where the Robin parameter a must be tuned (or optimized) for good performance. We show that the 2-Lagrange multiplier method can be analyzed using matrix analytical techniques and we produce sharp condition number estimates.

1 Introduction

Consider the model problem

$$-\Delta u = f \text{ in } \Omega \text{ and } u = 0 \text{ on } \partial\Omega, \tag{1}$$

where Ω is the domain, f is a given forcing and $u \in H_0^1(\Omega)$ is the unknown solution. In the present paper, we describe a symmetric 2-Lagrange multiplier (S2LM) domain decomposition method to solve elliptic problems such as (1). When we discretize (1) using e.g. piecewise linear finite elements, we obtain a linear system of the form

$$A\mathbf{u} = \mathbf{f}, \tag{2}$$

where $\mathbf{u} \in \mathbb{R}^n$ is the finite element coefficient vector of the approximation to the solution u of (1).

We now consider the domain decomposition [9] $\Omega = \Gamma \cup \Omega_1 \cup \ldots \cup \Omega_p$, where $\Omega_1, \ldots, \Omega_p$ are the (open, disjoint) "subdomains" and $\Gamma = \Omega \cap \bigcup_{k=1}^p \partial\Omega_k$ is the "artificial interface". We introduce the "local problems"

$$\begin{cases} -\Delta u_k = f & \text{in } \Omega_k, \quad \text{(PDE)} \\ u_k = 0 & \text{on } \partial\Omega_k \cap \partial\Omega, \quad \text{(natural b.c.)} \\ (a + D_\nu)u_k = \lambda_k & \text{on } \partial\Omega_k \cap \Gamma, \quad \text{(artificial b.c.)} \end{cases} \tag{3}$$

where $a > 0$ is the Robin tuning parameter and $k = 1, \ldots, p$ and D_ν denotes the directional derivative in the outwards pointing normal ν of $\partial\Omega_k$. The interface Γ is

R. Bank et al. (eds.), *Domain Decomposition Methods in Science and Engineering XX*, Lecture Notes in Computational Science and Engineering 91,
DOI 10.1007/978-3-642-35275-1_29,

artificial in that it is not a natural part of the "physical problem" (1) but instead is introduced purely for the purpose of calculation.

We again discretize the systems (3) using a finite element method. The Robin b.c. in (3) gives rise to a mass matrix on the interface $\Gamma \cap \partial\Omega_k$, which we lump. If the grid is uniform, this mass matrix is aI (we absorb any h factors into the a coefficient) – we make this simplification for the remainder of the present paper.

$$\begin{bmatrix} A_{Ilk} & A_{I\Gamma k} \\ A_{\Gamma Ik} & A_{\Gamma\Gamma k} + aI \end{bmatrix} \overbrace{\begin{bmatrix} \mathbf{u}_{Ik} \\ \mathbf{u}_{\Gamma k} \end{bmatrix}}^{\mathbf{u}_k} = \overbrace{\begin{bmatrix} \mathbf{f}_{Ik} \\ \mathbf{f}_{\Gamma k} \end{bmatrix}}^{\mathbf{f}_k} + \begin{bmatrix} 0 \\ \boldsymbol{\lambda}_k \end{bmatrix}. \tag{4}$$

Here, we have used the suggestive subscripts I for interior nodes and Γ for the artificial interface nodes.

The FETI-2LM algorithm was introduced in [4] for cases without cross-points, while the general case including cross points was introduced and analyzed in [7]. The method consists of finding the value of $\boldsymbol{\lambda} = [\boldsymbol{\lambda}_1^T, \ldots, \boldsymbol{\lambda}_p^T]^T$ which yields solutions $\mathbf{u}_1, \ldots, \mathbf{u}_p$ to (4) in such a way that $\mathbf{u}_1, \ldots, \mathbf{u}_p$ meet continuously across Γ and glue together into the unique solution $\mathbf{u}$ of (2).

The main result of the present paper is a new estimate of the condition number of FETI-2LM algorithms using matrix analytical techniques. This new idea produces sharp condition number estimates with much more straightforward proof techniques than the techniques used in [7] (where the estimates are not sharp). As a result, the present paper is a logical follow-up to [7].

The present paper focuses on 1-level algorithms which are known not to scale. Scalable algorithms are considered in [8] and [3].

Our paper is organized as follows. In Sect. 2, we give the symmetric 2-Lagrange multiplier method for general domains with cross points. In Sect. 3, we give spectral estimates including our main result, Theorem 1, on the condition number of the symmetric 2-Lagrange multiplier system. In Sect. 4, we verify this Theorem with some numerical experiments.

2 The Symmetric 2-Lagrange Multiplier Method

We now describe the 2-Lagrange multiplier method that we analyze in the present paper. Consider the local problems (4) and eliminate the interior degrees of freedom to obtain the relation

$$a \overbrace{\begin{bmatrix} \mathbf{u}_{\Gamma 1} \\ \vdots \\ \mathbf{u}_{\Gamma p} \end{bmatrix}}^{\mathbf{u}_G} = \overbrace{\begin{bmatrix} a(S_1 + aI)^{-1} & & \\ & \ddots & \\ & & a(S_p + aI)^{-1} \end{bmatrix}}^{Q} \left(\overbrace{\begin{bmatrix} \mathbf{g}_1 \\ \vdots \\ \mathbf{g}_p \end{bmatrix}}^{\mathbf{g}} + \overbrace{\begin{bmatrix} \boldsymbol{\lambda}_1 \\ \vdots \\ \boldsymbol{\lambda}_p \end{bmatrix}}^{\boldsymbol{\lambda}} \right), \tag{5}$$

where

$$S_k = A_{\Gamma\Gamma k} - A_{\Gamma Ik} A_{Ilk}^{-1} A_{I\Gamma k} \quad \text{and} \quad \mathbf{g}_k = \mathbf{f}_{\Gamma k} - A_{\Gamma Ik} A_{Ilk}^{-1} \mathbf{f}_{Ik}$$

are the "Dirichlet-to-Neumann maps" and "accumulated right-hand-sides" and where $\mathbf{u}_{\Gamma j}$ denotes those degrees of freedom of the local solution $\mathbf{u}_j$ associated with the artificial interface Γ.

The matrices S_k are symmetric and semidefinite. Since $Q = a(S+aI)^{-1}$, we find that the spectrum $\sigma(Q)$ is contained in the set $[\varepsilon, 1-\varepsilon] \cup \{1\}$ for some $\varepsilon > 0$. The eigenvalue 1 of Q comes from the kernel of S and hence the kernel of $Q-I$ is spanned by the indicating functions of the subdomains that "float".

2.1 Relations Between (4) and (2) and Continuity

We define the boolean restriction matrix R_k by selecting rows of the $n \times n$ identity matrix corresponding to those vertices of Ω that are in $\bar{\Omega}_k \cap \Omega$. As a result, from a finite element coefficient vector $\mathbf{v}$ corresponding to a finite element function $v \in H_0^1(\Omega)$, we can define a finite element coefficient vector $\mathbf{v}_k = R_k \mathbf{v}$, which corresponds to a finite element function $v \in H^1(\Omega_k) \cap H_0^1(\Omega)$, which is obtained by restricting v to Ω_k.

The identity $\int_\Omega = \sum_{k=1}^p \int_{\Omega_k}$ induces the following relations between (4) and (2):

$$A = \sum_{k=1}^{p} R_k^T \overbrace{\begin{bmatrix} A_{Ik} & A_{I\Gamma k} \\ A_{\Gamma Ik} & A_{\Gamma\Gamma k} \end{bmatrix}}^{A_{Nk}} R_k \quad \text{and} \quad \mathbf{f} = \sum_{k=1}^{p} R_k^T \mathbf{f}_k. \tag{6}$$

Each interface vertex $\mathbf{x}_i \in \Gamma$ is adjacent to $m_i \geq 2$ subdomains. As a result, the "many-sided trace" $\mathbf{u}_G$ defined by (5) contains m_i entries corresponding to $\mathbf{x}_i$, one per subdomain adjacent to $\mathbf{x}_i$. We define the orthogonal projection matrix K which averages function values for each interface vertex $\mathbf{x}_i$. A many-sided trace $\mathbf{u}_G$ corresponds to local functions $\mathbf{u}_1, \ldots, \mathbf{u}_p$ that meet continuously across Γ if and only if

$$K\mathbf{u}_G = \mathbf{u}_G. \tag{7}$$

2.2 A Problem in $\boldsymbol{\lambda}$

The **symmetric 2-Lagrange multiplier** (S2LM) system is given by

$$(Q-K)\boldsymbol{\lambda} = -Q\mathbf{g}. \tag{8}$$

We further let E be the orthogonal projection onto the kernel of $Q-I$.

Lemma 1. *Assume that $\|EK\| < 1$. The problem (2) is equivalent to (8).*

Proof. In order to solve (2) using local problems (4), one should find Robin boundary values $\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_p$ which result in local solutions $\mathbf{u}_1, \ldots, \mathbf{u}_p$ that meet continuously across Γ. As a result, we impose the condition (7), which we multiply by $a > 0$ and convert to an expression in $\boldsymbol{\lambda}$ using (5) to obtain $Ka(S+aI)^{-1}(\boldsymbol{\lambda}+\mathbf{g}) = a(S+aI)^{-1}(\boldsymbol{\lambda}+\mathbf{g})$ or

$$(I-K)Q\boldsymbol{\lambda} = (K-I)Q\mathbf{g} \tag{9}$$

With this continuity condition, there is clearly a unique $\mathbf{u}$ which restricts to the $\mathbf{u}_j$:

$$\mathbf{u}_j = R_j\mathbf{u}, \quad j = 1,\ldots,p. \tag{10}$$

Imposing continuity is not sufficient, we must also ensure that the "fluxes" match. Indeed, if we impose on the solution $\mathbf{u}$ of (10) that the Eq. (2) should hold, one obtains

$$\mathbf{f} = A\mathbf{u} \overset{(6)}{=} \sum_{j=1}^{p} R_j^T A_{Nj} R_j \mathbf{u} \overset{(10)}{=} \sum_{j=1}^{p} R_j^T A_{Nj} \mathbf{u}_j \tag{11}$$

$$\overset{(4),(6)}{=} \mathbf{f} + \sum_{j=1}^{p} R_j^T \begin{pmatrix} 0 \\ \boldsymbol{\lambda}_j - a\mathbf{u}_{\Gamma j} \end{pmatrix} \tag{12}$$

Canceling the $\mathbf{f}$ terms on each side and multiplying by K, we obtain $K\boldsymbol{\lambda} - Ka\mathbf{u}_G = 0$. Using (5), we obtain

$$K(Q-I)\boldsymbol{\lambda} = -KQ\mathbf{g}. \tag{13}$$

We add (9) and (13) to obtain (8).

To see that the solution of (8) is unique, observe that the ranges of E and K intersect trivially by the hypothesis that $\|EK\| < 1$. As a result, the eigenspace of Q of eigenvalue 1 intersects trivially with the range of K and $Q-K$ is nonsingular. □

We will further discuss the choice of the parameter a in Sect. 3.1.

3 Spectral Estimates

If we use GMRES or MINRES on the symmetric indefinite system (8), the residual norm can be estimated as a function of the condition number of $Q-K$, cf. [2]. In order to estimate the condition number of $Q-K$, we begin by giving a canonical form for the pair of projections E and K.

Lemma 2. *Let E and K be orthogonal projections. There is a choice of orthonormal basis that block diagonalizes E and K simultaneously and such that the blocks E_k and K_k of E and K satisfy*

$$E_k \in \left\{0, 1, \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}\right\} \quad \text{and} \quad K_k \in \left\{0, 1, \begin{bmatrix} c_k^2 & c_k s_k \\ c_k s_k & s_k^2 \end{bmatrix}\right\}, \tag{14}$$

where $c_k = \cos\theta_k > 0$, $s_k = \sin\theta_k > 0$ and $\theta_k \in (0, \pi/2)$ is a "principal angle" relating E and K.

The canonical form (14) can be obtained from the CS decomposition [1] by starting from $E = \mathrm{diag}(I, 0)$ and picking orthonormal bases for the range and kernel of K. Due to space constraints, we omit this argument.

We also give a technical lemma which describes the spectrum of a sum of certain symmetric matrices.

Lemma 3. *Let X, Y be symmetric matrices of dimensions $m \times m$. Let $0 < y_{\min} < y_{\max}$ and assume that $|\sigma(Y)| \subset [y_{\min}, y_{\max}]$. Denote by $\rho(X)$ the spectral radius of X and assume that $\rho(X) < y_{\min}$. Then,*

$$|\sigma(X+Y)| \subset [y_{\min} - \rho(X), y_{\max} + \rho(X)]. \tag{15}$$

Proof. This follows from a Theorem of Weyl [5, Theorem 4.3.1, pp. 181–182]. □

3.1 Condition Number of $Q - K$

We now come to our main result.

Theorem 1. *Let $\varepsilon > 0$. Assume that $\sigma(Q) \subset [\varepsilon, 1-\varepsilon] \cup \{1\}$. Let E, K be orthogonal projections and assume that $\|EK\| < 1$. Then we have the sharp estimates*

$$|\sigma(Q-K)| \subset \left[\frac{\varepsilon + \sqrt{(1+\varepsilon)^2 - 4\|EK\|^2\varepsilon} - 1}{2}, 1 \right], \quad \textit{and} \tag{16}$$

$$\kappa(Q-K) \leq \frac{2}{\varepsilon + \sqrt{(1+\varepsilon)^2 - 4\|EK\|^2\varepsilon} - 1} = O((1-\|EK\|)^{-1}\varepsilon^{-1}). \tag{17}$$

Proof. Let $X = Q - \frac{1}{2}I - \varepsilon E$ and $Y = \frac{1}{2}I + \varepsilon E - K$. Then, $Q - K = X + Y$ and we are in a position to use Lemma 3. We now estimate the spectral properties of X and Y.

Spectral properties of X: Recall that E projects onto the eigenspace of Q with eigenvalue 1. As a result, after some orthonormal change of basis, we find that $Q = \operatorname{diag}(Q_0, I)$ and $E = \operatorname{diag}(0, I)$ and hence

$$\rho(X) \leq \frac{1}{2} - \varepsilon. \tag{18}$$

Spectral properties of Y: Lemma 2 shows that E and K block diagonalize simultaneously and Y is also block diagonal in the same basis. Using (14), we find that the kth block Y_k of Y is given by

$$Y_k = \begin{cases} \frac{1}{2} & \text{if } E_k = K_k = 0, \\ -\frac{1}{2} & \text{if } E_k = 0,\ K_k = 1, \\ \frac{1}{2} + \varepsilon & \text{if } E_k = 1,\ K_k = 0, \\ \begin{bmatrix} \frac{1}{2} + \varepsilon - c_k^2 & -c_k s_k \\ -c_k s_k & \frac{1}{2} - s_k^2 \end{bmatrix} & \text{otherwise;} \end{cases} \tag{19}$$

where the case $E_k = K_k = 1$ is excluded by the hypothesis that $\|EK\| < 1$. As a result, the eigenvalues of Y_k are in the set $\{\pm\frac{1}{2}, \frac{1}{2} + \varepsilon, \lambda_\pm(c_k^2)\}$, where

$$\lambda_\pm(c_k^2) = \frac{\varepsilon \pm \sqrt{(1+\varepsilon)^2 - 4c_k^2\varepsilon}}{2}. \tag{20}$$

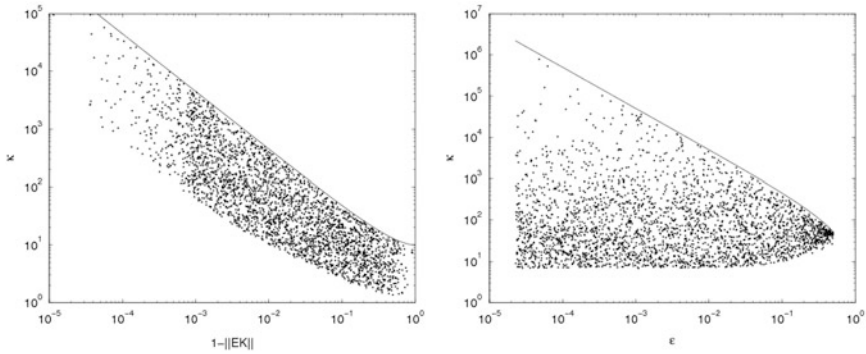


Fig. 1. Comparing random $Q-K$ (points) versus the estimate (17) (solid). *Left*: $\varepsilon = 0.1$, varying $\|EK\|$, 3,000 repetitions. *Right*: $\|EK\| = 0.99$, varying ε, 3,000 repetitions

Note that $\|EK\| = \sqrt{\rho(EKE)} = \max_k c_k$ and that the functions $\lambda_\pm(c_k^2)$ are monotonic in c_k^2. Hence, we find the following bounds for the modulus of an eigenvalue of Y:

$$|\sigma(Y)| \subset \left[\overbrace{\frac{\sqrt{(1+\varepsilon)^2 - 4\|EK\|^2\varepsilon} - \varepsilon}{2}}^{y_{\min}}, \overbrace{\frac{1}{2}+\varepsilon}^{y_{\max}}\right]. \tag{21}$$

Combining (15), (18), and (21) gives (16).

The examples $Q = \operatorname{diag}(1, 1-\varepsilon)$ and $K = \begin{bmatrix} c^2 & c\sqrt{1-c^2} \\ c\sqrt{1-c^2} & 1-c^2 \end{bmatrix}$ for $c = 0$ and $c = \|EK\|$ give the extreme eigenvalues of (21) and hence our estimates are sharp. □

In view of Theorem 1, the Robin parameter a should be chosen so as to make ε as large as possible. This occurs precisely when a is the geometric mean of the extremal positive eigenvalues of S. More details can be found in [7].

4 Numerical Verification

We verify numerically the validity of Theorem 1 by generating random 5×5 matrices Q and E as follows. We set $Q = \operatorname{diag}(\varepsilon, q, 1-\varepsilon, 1, 1)$ where q is chosen randomly between ε and $1-\varepsilon$. We generate randomly a 2-dimensional space and set K to be the orthogonal projection onto that space. We compare the resulting condition number $\kappa = \kappa(Q-K)$ against (17), cf. Fig. 1.

We observe that our estimates are correct and sharp for such "generic" random matrices, although some "lucky" random matrices produce much milder condition numbers than our estimates.

5 Conclusions

We have analyzed a domain decomposition method with optimized Robin boundary conditions. Our estimates rely on new matrix analytical techniques and are sharp. By further estimating the quantities $\|EK\|$ and ε (cf. [7]) our estimates are consistent with and generalize the estimates calculated using Fourier transforms in the optimized Schwarz literature (e.g. [6]). An upcoming paper [8] will further analyze the weak scaling property of a 2-level algorithm and large-scale implementations are being developed. There are also several remaining open problems, such as the analysis of FETI-2LM for nonsymmetric and/or nonlinear problems and the analysis of substructuring preconditioners.

Bibliography

[1] Chandler Davis and W. M. Kahan. Some new bounds on perturbation of subspaces. *Bulletin of the American Mathematical Society*, pages 863–868, 1969.

[2] Tobin A. Driscoll, Kim-Chuan Toh, and Lloyd N. Trefethen. From potential theory to matrix iterations in six steps. *SIAM Review*, pages 547–578, 1998.

[3] Stephen W. Drury and Sébastien Loisel. The performance of optimized Schwarz and 2-Lagrange multiplier preconditioners for GMRES. *Manuscript*, 2011.

[4] Charbel Farhat, Antonini Macedo, Michel Lesoinne, Francois-Xavier Roux, Frédéric Magoulès, and Armel de La Bourdonnaie. Two-level domain decomposition methods with lagrange multipliers for the fast iterative solution of acoustic scattering problems. *Computer Methods in Applied Mechanics and Engineering*, 184:213–239, 2000.

[5] Roger A. Horn and Charles R. Johnson. *Matrix Analysis*. Cambridge University Press, 1990.

[6] S. Loisel, J. Côté, M. J. Gander, L. Laayouni, and A. Qaddouri. Optimized domain decomposition methods for the spherical Laplacian. *SIAM Journal on Numerical Analysis*, 48:524–551, 2010.

[7] Sébastien Loisel. Condition number estimates for the nonoverlapping optimized Schwarz method and the 2-Lagrange multiplier method for general domains and cross points. *Submitted to SIAM Journal on Numerical Analysis*, 2011.

[8] Sébastien Loisel. Condition number estimates and weak scaling for 2-level 2-Lagrange multiplier methods for general domains and cross points. *Submitted*, 2011.

[9] Andrea Toselli and Olof B. Widlund. *Domain Decomposition Methods – Algorithms and Theory*, volume 34 of *Springer Series in Computational Mathematics*. Springer Berlin Heidelberg, 2005.

Time Domain Maxwell Equations Solved with Schwarz Waveform Relaxation Methods

Yves Courvoisier[1] and Martin J. Gander[2]

[1] University of Geneva, Section de mathématiques, Case Postale 64, 1211 Genève 4
yves.courvoisier@unige.ch
[2] University of Geneva, Section de mathématiques, Case Postale 64, 1211 Genève 4
martin.gander@unige.ch

1 Introduction

It is very natural to solve time dependent problems with Domain Decomposition Methods by using an implicit scheme for the time variable and then applying a classical iterative domain decomposition method at each time step. This is however not what the Schwarz Waveform Relaxation (SWR) methods do. The SWR methods are a combination of the Schwarz Domain Decomposition methods, see [10], and the Waveform Relaxation algorithm, see [7]. Combined, one obtains a new method which decomposes the domain into subdomains on which time dependent problems are solved. Iterations are then introduced, where communication between subdomains is done at artificial interfaces along the whole time window.

This new approach has been introduced by Bjørhus [1] for hyperbolic problems with Dirichlet boundary conditions and was analyzed for the heat equation by Gander and Stuart [5]. Giladi and Keller [6] analyzed this same approach applied to the advection diffusion equation with constant coefficients. For the wave equation and SWR see [3] in which they treat the one-dimensional case with overlapping subdomains and for the n-dimensional case [4], again with overlap. In this paper, we analyze for the first time the SWR algorithm applied to the time domain Maxwell equations.

2 Maxwell Equations and the Schwarz Waveform Relaxation Algorithm

The global domain Ω is decomposed into non overlapping subdomains $\tilde{\Omega}_i$. We denote by Ω_i the domain $\tilde{\Omega}_i$ enlarged by a band of width δ inside of Ω. The part of $\partial\Omega_i$ in $\tilde{\Omega}_j$ is denoted Γ_{ij}, i.e. $\Gamma_{ij} := \partial\Omega_i \cap \overline{\tilde{\Omega}}_j$. If Ω_i possesses a part of the boundary of the global domain Ω, we denote it by $\Gamma_{i0} := \partial\Omega_i \cap \partial\Omega$. The SWR algorithm with *characteristic transmission conditions* for the time domain Maxwell equations is given by

R. Bank et al. (eds.), *Domain Decomposition Methods in Science and Engineering XX*, 263
Lecture Notes in Computational Science and Engineering 91,
DOI 10.1007/978-3-642-35275-1_30,

$$\begin{cases} -\varepsilon\partial_t \mathbf{E}^{i,n} + \nabla\times\mathbf{H}^{i,n} - \sigma\mathbf{E}^{i,n} = \mathbf{J}, & \Omega_i\times(0,T), \\ \mu\partial_t \mathbf{H}^{i,n} + \nabla\times\mathbf{E}^{i,n} = 0, & \Omega_i\times(0,T), \\ \mathscr{B}_{\mathbf{n}_i}(\mathbf{E}^{i,n},\mathbf{H}^{i,n}) = 0, & \Gamma_{i0}\times(0,T), \\ (\mathbf{E}^{i,n},\mathbf{H}^{i,n})(\mathbf{x},0) = (\mathbf{E}_0,\mathbf{H}_0), & \Omega_i, \\ \mathscr{B}_{\mathbf{n}_i}(\mathbf{E}^{i,n},\mathbf{H}^{i,n}) = \mathscr{B}_{\mathbf{n}_i}(\mathbf{E}^{j,n-1},\mathbf{H}^{j,n-1}), & \Gamma_{ij}\times(0,T), \end{cases} \tag{1}$$

where ε is the electric permittivity, μ the magnetic permeability and σ the conductivity. The indices i and j, always different, range over the indices of all subdomains, i.e. $i,j\in\{1,2,\ldots,I\}$ with $i\neq j$ and I being the number of subdomains. In the algorithm $\mathbf{n}_i$ is the unit outward normal vector to Ω_i. The impedance

$$\mathscr{B}_{\mathbf{n}}(\mathbf{E},\mathbf{H}) := \frac{\mathbf{E}}{Z}\times\mathbf{n} + \mathbf{n}\times(\mathbf{H}\times\mathbf{n}),$$

plays the role of the Dirichlet value for this hyperbolic system [2] and corresponds to the inward characteristic variables of the Maxwell equations. The last line of (1), which is called the *characteristic transmission condition*, establishes how the subdomains communicate with each other.

3 Convergence in a Finite Number of Steps

From now on, we restrict our analysis to the specific situation where $\Omega=\mathbb{R}^3$ which is subdivided into two subdomains

$$\Omega_1 = (-\infty,L]\times\mathbb{R}^2, \quad \Omega_2 = [0,+\infty)\times\mathbb{R}^2. \tag{2}$$

The artificial boundaries are therefore given by $\Gamma_{12}=\{L\}\times\mathbb{R}^2$ and $\Gamma_{21}=\{0\}\times\mathbb{R}^2$ with an overlap of width L. We also choose the coefficients ε, μ and σ to be constant.

Maxwell equations describe the motion of electromagnetic waves which propagate at finite speed, namely the speed of light in the vacuum. This fact has been proven for a broad class of hyperbolic systems, see for instance [8]; the Maxwell equations are simply one such example. The speed of propagation is given by $c := 1/\sqrt{\varepsilon\mu}$, which is constant.

Remark 1. The next result also holds when the coefficients are non constant and with a domain Ω decomposed into many subdomains Ω_i having a more complicated geometry and non constant overlap width.

Proposition 1 (Convergence in a finite number of steps). *The SWR algorithm (1) for two subdomains defined in (2) with overlap L converges as soon as the number of iterations n satisfies*

$$n > \frac{Tc}{L},$$

where T is the length of the time interval and $c=1/\sqrt{\varepsilon\mu}$ *is the speed of propagation.*

Proof. The Maxwell equations are linear and thus allow us to restrict our attention to the error equations, i.e. (1) where $\mathbf{J} = 0$ and $(\mathbf{E}_0, \mathbf{H}_0) = 0$. We prove in the following that for $t < t_n := n\frac{L}{c}$,

$$\mathrm{Supp}(\mathbf{E}^{i,n+1}, \mathbf{H}^{i,n+1})(t) = \emptyset, \quad t < t_n. \tag{3}$$

The error of the Maxwell equations is non-zero at iteration one only because the initial guesses $(\mathbf{E}^{i,0}, \mathbf{H}^{i,0})$ are non-zero on the artificial boundaries Γ_{ij}. The speed of propagation is finite and thus the error propagates from the artificial boundaries inside the domain Ω_i. For the first iteration we have that

$$\mathrm{Supp}(\mathbf{E}^{i,1}, \mathbf{H}^{i,1})(t) \subset \{\mathbf{x} \in \Omega_i | \mathrm{dist}(x, \Gamma_{ij}) < tc, j \neq i, j \in \{1,2\}\},$$

since after a time t, the electromagnetic wave can only have propagated on a distance tc from the artificial boundaries. The overlap is of width L, hence $(\mathbf{E}^{1,1}, \mathbf{H}^{1,1})(0, y, z, t)$ and $(\mathbf{E}^{2,1}, \mathbf{H}^{2,1})(L, y, z, t)$ are zero unless $tc > L$, i.e. unless the time is greater or equal to $t_1 := \frac{L}{c}$.

For the next iteration we have that the trace of $(\mathbf{E}^{1,1}, \mathbf{H}^{1,1})$ at Γ_{21} and $(\mathbf{E}^{2,1}, \mathbf{H}^{2,1})$ at Γ_{12} are zero for times $t < t_1$, i.e. $B_{\mathbf{n}_i}(\mathbf{E}^{j,n-1}, \mathbf{H}^{j,n-1}) = 0$ at Γ_{ij} for $n = 2$ and $t < t_1$. Therefore, when solving for $(\mathbf{E}^{i,2}, \mathbf{H}^{i,2})$ we see that for $t < t_1$, we have zero boundary conditions and zero initial condition, hence

$$(\mathbf{E}^{i,2}, \mathbf{H}^{i,2})(\mathbf{x}, t) = 0, \quad \text{for } t < t_1.$$

For times $t > t_1$, we have a similar result as for the first iteration, namely

$$\mathrm{Supp}(\mathbf{E}^{i,2}, \mathbf{H}^{i,2})(t) \subset \{\mathbf{x} \in \Omega_i | \mathrm{dist}(x, \Gamma_{ij}) < (t - t_1)c, j \neq i, j \in \{1,2\}\}.$$

We define $t_2 := \frac{L}{c} + t_1 = 2t_1$, such that $\mathrm{Supp}(\mathbf{E}^{i,2}, \mathbf{H}^{i,2})(t) = 0$ on Γ_{ji} for $t < t_2$. And so forth for the following iterations, which proves (3).

Hence, if T, the length of the time window, is finite and $t_n := n\frac{L}{c} > T$, the solution $(\mathbf{E}^{i,n+1}, \mathbf{H}^{i,n+1})$ is zero and the algorithm has converged.

4 Convergence of the SWR Algorithm

Under the same setting (2) as in previous section, we prove that the SWR algorithm (1) also has a contraction factor.

Theorem 1. *The convergence factor of the classical Schwarz Waveform Relaxation algorithm (1) in the frequency domain with domain decomposition (2) is given by*

$$\rho(s, k_y, k_z, L, \sigma) = \left| \frac{\sqrt{|\mathbf{k}|^2 + \mu s^2 \varepsilon + \mu s \sigma} - s\sqrt{\mu\varepsilon}}{\sqrt{|\mathbf{k}|^2 + \mu s^2 \varepsilon + \mu s \sigma} + s\sqrt{\mu\varepsilon}} e^{-L\sqrt{|\mathbf{k}|^2 + \mu s^2 \varepsilon + \mu s \sigma}} \right|,$$

where s is the Laplace variable, $\Re(s) \geq 0$, and $|\mathbf{k}|^2 = k_y^2 + k_z^2$ is the sum of the squares of the Fourier frequencies in the y and z directions.

Proof. We consider the error equations for which $\mathbf{J}$ and the initial condition are zero. We first apply the Laplace transform to (1) which transforms the time t into a complex frequency s with $\Re(s) \geq 0$ and transforms the derivative with respect to t into a multiplication by s. Then we apply a Fourier transform in the y and z directions and obtain,

$$\frac{\partial}{\partial x}\begin{bmatrix}\check{E}_2\\ \check{E}_3\\ \check{H}_2\\ \check{H}_3\end{bmatrix} + \begin{bmatrix} 0 & 0 & -\frac{k_y k_z}{\varepsilon s+\sigma} & \frac{k_y^2}{\varepsilon s+\sigma}+\mu s\\ 0 & 0 & -\frac{k_z^2}{\varepsilon s+\sigma}-\mu s & \frac{k_y k_z}{\varepsilon s+\sigma}\\ \frac{k_y k_z}{\mu s} & -\frac{k_y^2}{\mu s}-(\varepsilon s+\sigma) & 0 & 0\\ \frac{k_z^2}{\mu s}+\varepsilon s+\sigma & -\frac{k_y k_z}{\mu s} & 0 & 0\end{bmatrix}\begin{bmatrix}\check{E}_2\\ \check{E}_3\\ \check{H}_2\\ \check{H}_3\end{bmatrix} = 0 \quad (4)$$

For components $\check{E}_1$ and $\check{H}_1$, we have two algebraic equations

$$\begin{aligned} -\varepsilon s\check{E}_1 + ik_y\check{H}_3 - ik_z\check{H}_2 - \sigma\check{E}_1 = 0,\\ \mu s\check{H}_1 + ik_y\check{E}_3 - ik_z\check{E}_2 = 0.\end{aligned}$$

The solution of (4) is given by a linear combination of the eigenvectors times an exponential of the corresponding eigenvalue,

$$\begin{aligned}(\check{E}_2^{1,n},\check{E}_3^{1,n},\check{H}_2^{1,n},\check{H}_3^{1,n})^T &= (\alpha_1^n\mathbf{v}_1+\alpha_2^n\mathbf{v}_2)e^{-\lambda(x-L)} + (\alpha_3^n\mathbf{v}_3+\alpha_4^n\mathbf{v}_4)e^{\lambda(x-L)},\\ (\check{E}_2^{2,n},\check{E}_3^{2,n},\check{H}_2^{2,n},\check{H}_3^{2,n})^T &= (\beta_1^n\mathbf{v}_1+\beta_2^n\mathbf{v}_2)e^{-\lambda x} + (\beta_3^n\mathbf{v}_3+\beta_4^n\mathbf{v}_4)e^{\lambda x}.\end{aligned} \quad (5)$$

where $\lambda = \sqrt{|k|^2+\mu s^2\varepsilon+\mu s\sigma}$ and the eigenvalues are $\lambda_{1,2} = -\lambda$ and $\lambda_{3,4} = \lambda$. The corresponding eigenvectors are

$$\begin{aligned}\mathbf{v}_1 &= \begin{pmatrix}\frac{k_y k_z}{\lambda(\varepsilon s+\sigma)}\\ \frac{k_z^2+\mu s^2\varepsilon+\mu s\sigma}{\lambda(\varepsilon s+\sigma)}\\ 1\\ 0\end{pmatrix}, \mathbf{v}_2 = \begin{pmatrix}-\frac{k_y^2+\mu s^2\varepsilon+\mu s\sigma}{\lambda(\varepsilon s+\sigma)}\\ -\frac{k_y k_z}{\lambda(\varepsilon s+\sigma)}\\ 0\\ 1\end{pmatrix},\\ \mathbf{v}_3 &= \begin{pmatrix}-\frac{k_y k_z}{\lambda(\varepsilon s+\sigma)}\\ -\frac{k_z^2+\mu s^2\varepsilon+\mu s\sigma}{\lambda(\varepsilon s+\sigma)}\\ 1\\ 0\end{pmatrix}, \mathbf{v}_4 = \begin{pmatrix}\frac{k_y^2+\mu s^2\varepsilon+\mu s\sigma}{\lambda(\varepsilon s+\sigma)}\\ \frac{k_y k_z}{\lambda(\varepsilon s+\sigma)}\\ 0\\ 1\end{pmatrix}.\end{aligned} \quad (6)$$

The speed of propagation is finite. The wave of the error equations propagates starting from the interfaces. Therefore, no wave is coming from the infinite boundary and then the growing exponential term of (5) is not present in the solution, i.e. $\alpha_1 = \alpha_2 = \beta_3 = \beta_4 = 0$. Hence,

$$\begin{aligned}(\check{E}_2^{1,n},\check{E}_3^{1,n},\check{H}_2^{1,n},\check{H}_3^{1,n})^T &= (\alpha_3^n\mathbf{v}_3+\alpha_4^n\mathbf{v}_4)e^{\lambda(x-L)},\\ (\check{E}_2^{2,n},\check{E}_3^{2,n},\check{H}_2^{2,n},\check{H}_3^{2,n})^T &= (\beta_1^n\mathbf{v}_1+\beta_2^n\mathbf{v}_2)e^{-\lambda x}.\end{aligned} \quad (7)$$

To determine the values of α_i and β_i, we need to use the transmission conditions. They are, for the first subdomain, $\mathscr{B}_{\mathbf{n}}(\check{\mathbf{E}}^{1,n},\check{\mathbf{H}}^{1,n}) = \mathscr{B}_{\mathbf{n}}(\check{\mathbf{E}}^{2,n-1},\check{\mathbf{H}}^{2,n-1})$ with $\mathbf{n} = (1,0,0)^T$, i.e.

$$\begin{bmatrix} \frac{1}{Z}\check{E}_3^{1,n} + \check{H}_2^{1,n} \\ -\frac{1}{Z}\check{E}_2^{1,n} + \check{H}_3^{1,n} \end{bmatrix} = \begin{bmatrix} \frac{1}{Z}\check{E}_3^{2,n-1} + \check{H}_2^{2,n-1} \\ -\frac{1}{Z}\check{E}_2^{2,n-1} + \check{H}_3^{2,n-1}. \end{bmatrix}$$

We substitute the values of the electric and magnetic fields by their values given in (7). This gives an equation relating $\boldsymbol{\alpha}^n = (\alpha_3^n, \alpha_4^n)^T$ and $\boldsymbol{\beta}^n = (\beta_1^n, \beta_2^n)^T$,

$$A_1 \boldsymbol{\alpha}^n = A_2 e^{-\lambda L} \boldsymbol{\beta}^{n-1}, \tag{8}$$

where matrices A_1 and A_2 are given by

$$\begin{aligned} A_1 &= \begin{bmatrix} -(k_z^2 + \mu s^2\varepsilon + \mu s\sigma) + Z\lambda(\varepsilon s + \sigma) & k_y k_z \\ k_y k_z & -(k_y^2 + \mu s^2\varepsilon + \mu s\sigma) + Z\lambda(\varepsilon s + \sigma) \end{bmatrix}, \\ A_2 &= \begin{bmatrix} k_z^2 + \mu s^2\varepsilon + \mu s\sigma + Z\lambda(\varepsilon s + \sigma) & -k_y k_z \\ -k_y k_z & k_y^2 + \mu s^2\varepsilon + \mu s\sigma + Z\lambda(\varepsilon s + \sigma) \end{bmatrix}. \end{aligned} \tag{9}$$

We do the same computations for the second subdomain for which we have the transmission conditions $\mathscr{B}_{-\mathbf{n}}(\hat{\mathbf{E}}^{2,n},\hat{\mathbf{H}}^{2,n}) = \mathscr{B}_{-\mathbf{n}}(\hat{\mathbf{E}}^{1,n-1},\hat{\mathbf{H}}^{1,n-1})$, and obtain

$$A_1 \boldsymbol{\beta}^n = A_2 e^{-\lambda L} \boldsymbol{\alpha}^{n-1}. \tag{10}$$

We isolate $\boldsymbol{\alpha}^n$ and $\boldsymbol{\beta}^n$ in (8) and (10) and iterate one more time to obtain

$$\boldsymbol{\alpha}^n = (A_1^{-1}A_2)^2 e^{-2\lambda L} \boldsymbol{\alpha}^{n-2}, \quad \boldsymbol{\beta}^n = (A_1^{-1}A_2)^2 e^{-2\lambda L} \boldsymbol{\beta}^{n-2}. \tag{11}$$

The parameters $\boldsymbol{\alpha}^n$ and $\boldsymbol{\beta}^n$ characterize completely the solution of (4), therefore the effective contraction factor after two iterations is given by the spectral radius of $(A_1^{-1}A_2)^2 e^{-2\lambda L}$. This matrix has eigenvalues

$$\nu_1 := \left(\frac{\lambda - s\sqrt{\varepsilon\mu}}{\lambda + s\sqrt{\varepsilon\mu}}\right)^2 e^{-2\lambda L}, \quad \nu_2 := \left(\frac{\lambda - s\sqrt{\varepsilon\mu} - Z\sigma}{\lambda + s\sqrt{\varepsilon\mu} + Z\sigma}\right)^2 e^{-2\lambda L}.$$

The largest eigenvalue in modulus is given by the first one which concludes the proof.

Corollary 1. *The SWR algorithm (1) with non-zero conductivity, $\sigma > 0$, converges in the L^2 norm, i.e. if we denote by $e^{i,n} := (E_2^{i,n}, E_3^{i,n}, H_2^{i,n}, H_3^{i,n})$, then*

$$||e^{i,n}(\Gamma_{ij}, t)||_2 \longrightarrow 0 \quad (n \to +\infty),$$

where Γ_{ij} is defined in (2) and $||\cdot||_2$ denotes the norm in $L^2(0,T;L^2(\mathbb{R}^2))$.

Proof. We use the notation $\check{e}^{i,n} = (\check{E}_2^{i,n}, \check{E}_3^{i,n}, \check{H}_2^{i,n}, \check{H}_3^{i,n})$ for the solution in the Fourier Laplace variables. From relations (11) with the notation $R := A_1^{-1}A_2 e^{-\lambda L}$ and iterating $2n$ times we obtain

$$\boldsymbol{\alpha}^{2n} = R^{2n}\boldsymbol{\alpha}^0, \quad \boldsymbol{\beta}^{2n} = R^{2n}\boldsymbol{\beta}^0.$$

The matrix R has eigenvalues ν_1 and ν_2 and therefore can be diagonalized using the matrix of eigenvectors S, i.e. $D = S^{-1}RS$. The following argument, for the first subdomain Ω_1, is similar also for the second one.

We define $\boldsymbol{\gamma}^n := S^{-1}\boldsymbol{\alpha}^n$ for all $n = 0, 1, \ldots$, and from (7) we can reconstruct the solution of $\check{e}^{1,2n}$ from the initial iterate,

$$\begin{aligned}\check{e}^{1,2n}(x,k_y,k_z,s) = e^{\lambda(x-L)}[\mathbf{v}_3\ \mathbf{v}_4]R^{2n}\boldsymbol{\alpha}^0 &= e^{\lambda(x-L)}[\mathbf{v}_3\ \mathbf{v}_4]SS^{-1}R^{2n}S\boldsymbol{\gamma}^0 \\ &= e^{\lambda(x-L)}[\mathbf{v}_3\ \mathbf{v}_4]SD^{2n}\boldsymbol{\gamma}^0.\end{aligned}$$

The diagonal matrix is of the form $D = \mathrm{diag}(\nu_1, \nu_2)$, hence we obtain a new form for the solution evaluated at $x = L$,

$$\check{e}^{1,2n}(L,k_y,k_z,s) = \nu_1^{2n}\gamma_1^0\mathbf{w}_1 + \nu_2^{2n}\gamma_2^0\mathbf{w}_2, \tag{12}$$

where $[\mathbf{w}_1\ \mathbf{w}_2] := [\mathbf{v}_3\ \mathbf{v}_4]S$.

Finally Theorem 7.23 of [9] shows that the limit $\check{e}^{i,n}(L,k_y,k_z,s)$ when $s = \xi + i\omega \to i\omega$ is the Fourier transform of $e^{i,n}$ in the y, z and t variables. Therefore the Plancherel theorem applies and

$$||e^{i,n}(L,y,z,t)||_2 = ||\check{e}^{i,n}(L,k_y,k_z,i\omega)||_2,$$

which implies by (12)

$$||e^{i,n}(L,y,z,t)||_2 = ||\nu_1^{2n}\gamma_1^0\mathbf{w}_1 + \nu_2^{2n}\gamma_2^0\mathbf{w}_2||_2$$

By the dominated convergence theorem we can insert the limit, when n goes to infinity, into the norm and, since $\lim_{n\to\infty}\nu_i$ is almost everywhere zero for $i = 1, 2$, it concludes the proof.

5 Numerical Experiments

For this section we restrict the geometry of the global domain to $\Omega = [0,1]^3$ and to subdomains

$$\Omega_1 = [0, \frac{1}{2} + 2\Delta x] \times [0,1] \times [0,1], \quad \Omega_2 = [\frac{1}{2}, 1] \times [0,1] \times [0,1],$$

where Δx is the spatial mesh size in the direction x. We consider a time window of length $T = 1$. The parameters ε, μ and σ are constant and equal to one. On the physical domain we set boundary conditions for perfectly conducting medium.

The discretization is done with the Yee scheme which is explicit in time. We set a global grid on the whole domain Ω having 24 grid points in each direction x, y and z. The overlap is of 2 mesh points. The number of grid points for the time variable is $N = 144$ which guarantees that the CFL condition is satisfied. Since the

domain is bounded, only a finite number of discrete frequencies are possible. Since the domain is of width one, the minimum frequency in space is given by $k_{min} = \pi$ and the maximum by $k_{max} = \frac{\pi}{\Delta y}$. Equivalently for the time frequencies we have $\omega_{max} = \frac{\pi}{\Delta t}$. Since there is no finite value imposed, we take $\omega_{min} = \frac{\pi}{2T} = \frac{\pi}{2}$. The discrete frequencies are therefore given by

$$k_y, k_z \in \{\pi, 2\pi, \ldots, \frac{\pi}{\Delta y}\}, \quad \omega \in \{\frac{\pi}{2}, \pi, \ldots, \frac{\pi}{\Delta t}\}.$$

From Corollary 1 we have that

$$||e^{i,n}(L, y, z, t)||_2 \leq C \max_{(k_y, k_z, \omega)} |\nu_1|^n, \tag{13}$$

where the constant C is the maximum over all frequencies of $||\gamma_1^0 \mathbf{w}_1 + \frac{\nu_2}{\nu_1}\gamma_2^0 \mathbf{w}_2||_2$. We also expect the solution to converge in a finite number of iterations as shown in Fig. 1.

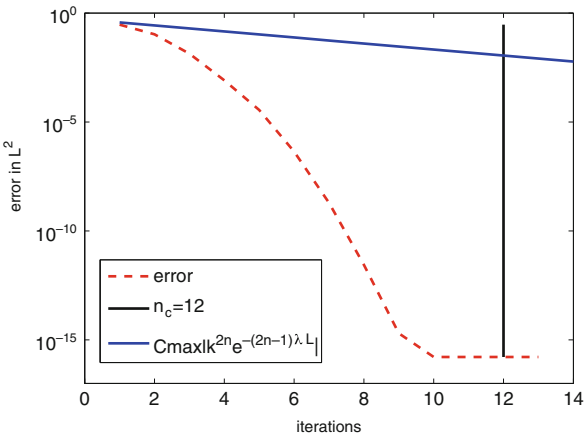


Fig. 1. The *plain blue line* is the *upper bound* in (13), and the *dashed line* is the error $||E_2^{1,n}||$ in the L^2 norm evaluated at the interface $x = b$ with respect to the iterations. The error converges before the relation of Proposition 1 is satisfied (*vertical line*)

Bibliography

[1] Morten Bjørhus. *On Domain Decomposition, Subdomain Iteration and Waveform Relaxation*. PhD thesis, University of Trondheim, Norway, 1995.

[2] V. Dolean, M.J. Gander, and L. Gerardo-Giorda. Optimized Schwarz methods for Maxwell's equations. *SIAM Journal on Scientific Computing*, 31(3):2193–2213, 2009.

[3] Martin J. Gander and Laurence Halpern. Méthodes de décomposition de domaines pour l'équation des ondes en dimension 1. *C. R. Acad. Sci. Paris Sér. I Math.*, 333(6):589–592, 2001.

[4] Martin J. Gander and Laurence Halpern. Absorbing boundary conditions for the wave equation and parallel computing. *Math. Comp.*, 74(249):153–176 (electronic), 2005.

[5] Martin J. Gander and Andrew M. Stuart. Space-time continuous analysis of waveform relaxation for the heat equation. *SIAM J. Sci. Comput.*, 19(6):2014–2031, 1998.

[6] Eldar Giladi and Herbert B. Keller. Space-time domain decomposition for parabolic problems. *Numer. Math.*, 93(2):279–313, 2002.

[7] Ekachai Lelarasmee, Albert E. Ruehli, and Alberto L. Sangiovanni-Vincentelli. The waveform relaxation method for time-domain analysis of large scale integrated circuits. *IEEE Transaction on Computer-aided Design of Integrated Cicuits and Systems*, CAD-1(3):131–145, 1982.

[8] Jeffrey Rauch. Precise finite speed with bare hands. *Methods Appl. Anal.*, 12 (3):267–277, 2005.

[9] Walter Rudin. *Functional analysis*. McGraw-Hill series in higher mathematics. McGraw-Hill Inc., 1973.

[10] H. A. Schwarz. Über einen Grenzübergang durch alternierendes Verfahren. *Vierteljahrsschrift der Naturforschenden Gesellschaft in Zürich*, 15:272–286, 1870.

Comparison of a One and Two Parameter Family of Transmission Conditions for Maxwell's Equations with Damping

M. El Bouajaji[1], V. Dolean[2], M. J. Gander[3] and S. Lanteri[1]

[1] NACHOS project-team, INRIA Sophia Antipolis - Méditerranée research center, F-06902 Sophia Antipolis Cedex, France Mohamed.El_bouajaji@inria.fr, Stephane.Lanteri@inria.fr
[2] Laboratoire J.A. Dieudonné, CNRS UMR 6621, F-06108 Nice Cedex, France dolean@unice.fr
[3] Mathematics Section, University of Geneva, CH-1211, Geneva, Switzerland martin.gander@unige.ch

1 Introduction

Transmission conditions between subdomains have a substantial influence on the convergence of iterative domain decomposition algorithms. For Maxwell's equations, transmission conditions which lead to rapidly converging algorithms have been developed both for the curl-curl formulation of Maxwell's equation, see [1–3], and also for first order formulations, see [6, 7]. These methods have well found their way into applications, see for example [9] and the references therein. It turns out that good transmission conditions are approximations of transparent boundary conditions. For each form of approximation chosen, one can try to find the best remaining free parameters in the approximation by solving a min-max problem. Usually allowing more free parameters leads to a substantially better solution of the min-max problem, and thus to a much better algorithm. For a particular one parameter family of transmission conditions analyzed in [4], we investigate in this paper a two parameter counterpart. The analysis, which is substantially more complicated than in the one parameter case, reveals that in one particular asymptotic regime there is only negligible improvement possible using two parameters, compared to the one parameter results. This analysis settles an important open question for this family of transmission conditions, and also suggests a direction for systematically reducing the number of parameters in other optimized transmission conditions.

2 Schwarz Methods for Maxwell's Equations

We consider in this paper a boundary value problem associated to three time-harmonic Maxwell equations with an impedance condition on the boundary of the

R. Bank et al. (eds.), *Domain Decomposition Methods in Science and Engineering XX*, Lecture Notes in Computational Science and Engineering 91,
DOI 10.1007/978-3-642-35275-1_31,

computational domain Ω,

$$\begin{aligned} -i\omega\varepsilon\mathbf{E} + \operatorname{curl}\mathbf{H} - \sigma\mathbf{E} = \mathbf{J},\ i\omega\mu\mathbf{H} + \operatorname{curl}\mathbf{E} = \mathbf{0},\ \Omega \\ \mathscr{B}_{\mathbf{n}}(\mathbf{E},\mathbf{H}) := \mathbf{n}\times\tfrac{\mathbf{E}}{Z} + \mathbf{n}\times(\mathbf{H}\times\mathbf{n}) = \mathbf{s},\ \partial\Omega. \end{aligned} \tag{1}$$

with $\mathbf{E},\mathbf{H}$ being the unknown electric and magnetic fields and ε,μ,σ being respectively the electric permittivity, magnetic permeability and the conductivity of the propagation medium and $\mathbf{n}$ the outward normal to $\partial\Omega$.
A family of Schwarz methods for (1) with a possibly non-overlapping decomposition of the domain Ω into Ω_1 and Ω_2, with interfaces $\Gamma_{12} := \partial\Omega_1 \cap \Omega_2$ and $\Gamma_{21} := \partial\Omega_2 \cap \Omega_1$, is given by

$$\begin{aligned} -i\omega\varepsilon\mathbf{E}^{1,n} + \operatorname{curl}\mathbf{H}^{1,n} - \sigma\mathbf{E}^{1,n} &= \mathbf{J} && \text{in } \Omega_1, \\ i\omega\mu\mathbf{H}^{1,n} + \operatorname{curl}\mathbf{E}^{1,n} &= \mathbf{0} && \text{in } \Omega_1, \\ (\mathscr{B}_{\mathbf{n}_1} + \mathscr{S}_1\mathscr{B}_{\mathbf{n}_2})(\mathbf{E}^{1,n},\mathbf{H}^{1,n}) &= (\mathscr{B}_{\mathbf{n}_1} + \mathscr{S}_1\mathscr{B}_{\mathbf{n}_2})(\mathbf{E}^{2,n-1},\mathbf{H}^{2,n-1}) && \text{on } \Gamma_{12}, \\ -i\omega\varepsilon\mathbf{E}^{2,n} + \operatorname{curl}\mathbf{H}^{2,n} - \sigma\mathbf{E}^{2,n} &= \mathbf{J} && \text{in } \Omega_2, \\ i\omega\mu\mathbf{H}^{2,n} + \operatorname{curl}\mathbf{E}^{2,n} &= \mathbf{0} && \text{in } \Omega_2, \\ (\mathscr{B}_{\mathbf{n}_2} + \mathscr{S}_2\mathscr{B}_{\mathbf{n}_1})(\mathbf{E}^{2,n},\mathbf{H}^{2,n}) &= (\mathscr{B}_{\mathbf{n}_2} + \mathscr{S}_2\mathscr{B}_{\mathbf{n}_1})(\mathbf{E}^{1,n-1},\mathbf{H}^{1,n-1}) && \text{on } \Gamma_{21}, \end{aligned} \tag{2}$$

where $\mathscr{S}_j$, $j = 1,2$ are tangential operators. For the case of constant coefficients and the domain $\Omega = \mathbb{R}^2$, with the Silver-Müller radiation condition $\lim_{r\to\infty} r\,(\mathbf{H}\times\mathbf{n} - \mathbf{E}) = 0$ and the two subdomains $\Omega_1 = (0,\infty)\times\mathbb{R}$, $\Omega_2 = (-\infty, L)\times\mathbb{R}$, $L \geq 0$, the following convergence result was obtained in [4] using Fourier analysis:

Theorem 1. *For $\sigma > 0$, if $\mathscr{S}_j$, $j = 1,2$ have the constant Fourier symbol*

$$\sigma_j = \mathscr{F}(\mathscr{S}_j) = -\frac{s - i\tilde{\omega}}{s + i\tilde{\omega}}, \quad \tilde{\omega} = \omega\sqrt{\varepsilon\mu}, \qquad s \in \mathbb{C}, \tag{3}$$

then the optimized Schwarz method (2), has the convergence factor

$$\rho(k,\tilde{\omega},Z,\sigma,L,s) = \left| \left(\frac{\sqrt{k^2 - \tilde{\omega}^2 + i\tilde{\omega}\sigma Z} - s}{\sqrt{k^2 - \tilde{\omega}^2 + i\tilde{\omega}\sigma Z} + s} \right) e^{-\sqrt{k^2 - \tilde{\omega}^2 + i\tilde{\omega}\sigma Z}L} \right|. \tag{4}$$

In order to obtain the most efficient algorithm, we choose σ_j, $j = 1,2$ such that ρ is minimal over the range of numerical frequencies $k \in K = [k_{\min}, k_{\max}]$, e.g. $k_{\min} = 0$ and $k_{\max} = \frac{C}{h}$ with h the mesh size and C a constant. We look for s of the form $s = p + iq$, such that (p,q) is solution of the min-max problem

$$\rho^* := \min_{p,q\geq 0} \left(\max_{k\in K} \rho(k,\tilde{\omega},Z,\sigma,L,p+iq) \right). \tag{5}$$

In [4] we have solved this min-max problem for the case $p = q$ without overlap, and we have obtained the following result:

Theorem 2. *For $\sigma > 0$ and $L = 0$, the solution of the min-max problem (5) with $p = q$ is for h small given by*

$$p^* = \frac{(\omega\sigma\mu)^{\frac{1}{4}}\sqrt{C}}{2^{\frac{1}{4}}\sqrt{h}} \quad \text{and} \quad \rho_1^* = 1 - \frac{2^{\frac{3}{4}}(\omega\sigma\mu)^{\frac{1}{4}}\sqrt{h}}{\sqrt{C}} + O(h). \tag{6}$$

For the overlapping case, we obtained in [8]:

Theorem 3. *For $\sigma > 0$ and $L = h$, a local minimum of the min-max problem (5) with $p = q$ is for h small given by*

$$p^* = \frac{(2\omega\sigma\mu)^{\frac{1}{3}}}{2h^{\frac{1}{3}}} \quad \text{and} \quad \rho_{1L}^* = 1 - 2^{\frac{7}{6}}(\omega\sigma\mu)^{\frac{1}{6}} h^{\frac{1}{3}} + O(h^{\frac{2}{3}}). \tag{7}$$

3 Analysis of the Two Parameter Family of Transmission Conditions

As before, we set $k_{\min} = 0$, $k_{\max} = \frac{C}{h}$ and denote by (p^*, q^*) a local minimum of (5). We first consider the non-overlapping case.

Theorem 4. *For $\sigma > 0$ and $L = 0$, a local minimum (p^*, q^*) of (5) is for h small given by*

$$p^* = \frac{3^{\frac{3}{8}}(\omega\sigma\mu)^{\frac{1}{4}}\sqrt{C}}{2^{\frac{3}{4}}\sqrt{h}}, q^* = \frac{3^{\frac{7}{8}}(2\omega\sigma\mu)^{\frac{1}{4}}\sqrt{C}}{6\sqrt{h}}, \rho_2^* = 1 - \frac{3^{\frac{3}{8}}(2\omega\sigma\mu)^{\frac{1}{4}}\sqrt{h}}{\sqrt{C}} + O(h). \tag{8}$$

Proof. By solving the min-max problem (5) numerically for different parameter values and different mesh sizes h, we observe that the solution of (5) equioscillates once, i.e. (p^*, q^*) is solution of

$$\rho(\bar{k}, \tilde{\omega}, \sigma, Z, 0, p^* + iq^*) = \rho(k_{\max}, \tilde{\omega}, \sigma, Z, 0, p^* + iq^*), \tag{9}$$

where $\bar{k}$ is an interior local maximum of ρ. We also observe the asymptotic behavior

$$\bar{k} \sim \bar{C}, \quad p^* \sim C_p h^{-\frac{1}{2}}, \quad q^* \sim C_q h^{-\frac{1}{2}}.$$

In order to determine the constants $\bar{C}$, C_p and C_q, it is necessary to have three equations. The first is (9), the second describes the interior local maximum of ρ in k,

$$\frac{\partial \rho}{\partial k}(\bar{k}, \tilde{\omega}, \sigma, Z, 0, p^* + iq^*) = 0,$$

and the third is the necessary condition for a local minimum of the min-max problem,

$$\begin{aligned}&\tfrac{d\rho}{dq}(k_{\max}, \tilde{\omega}, \sigma, Z, 0, p^* + iq^*) = \\ &\quad \tfrac{\partial \rho}{\partial q}(k_{\max}, \tilde{\omega}, \sigma, Z, 0, p^* + iq^*) + \tfrac{\partial \rho}{\partial p}(k_{\max}, \tilde{\omega}, \sigma, Z, 0, p^* + iq^*)\tfrac{\partial p}{\partial q} = 0.\end{aligned}$$

Since $\frac{d\rho}{dq}(k_{\max}, \tilde{\omega}, \sigma, Z, 0, p^* + iq^*) = \frac{d\rho}{dq}(\bar{k}, \tilde{\omega}, \sigma, Z, 0, p^* + iq^*)$ a similar expansion together with the previous one, gives

$$\frac{\partial p}{\partial q} = -\frac{\frac{\partial \rho}{\partial q}(k_{\max},\tilde{\omega},\sigma,Z,0,p^*+iq^*) - \frac{\partial \rho}{\partial q}(\bar{k},\tilde{\omega},\sigma,Z,0,p^*+iq^*)}{\frac{\partial \rho}{\partial p}(k_{\max},\tilde{\omega},\sigma,Z,0,p^*+iq^*) - \frac{\partial \rho}{\partial p}(\bar{k},\tilde{\omega},\sigma,Z,0,p^*+iq^*)},$$

and thus asymptotically, the three equations lead to the system

$$\begin{aligned}(\sqrt{A_1}+\bar{C}^2-\tilde{\omega}^2)(AC_p+BC_q)-2\sqrt{A_1}BC_q &= 0,\\ 2C_p(C_p^2+C_q^2)-C(BC_p+AC_q) &= 0,\\ A(C_q^2-C_p^2)+2C_pC_qB &= 0,\end{aligned}$$

where $A=\sqrt{2\sqrt{A_1}-A_2}$, $B=\sqrt{2\sqrt{A_1}+A_2}$, $A_1=\bar{C}^4-2(\bar{C}\tilde{\omega})^2+\tilde{\omega}^4+(\tilde{\omega}\sigma Z)^2$ and $A_2=2(\bar{C}^2-\tilde{\omega}^2)$. The solution of this system is

$$\bar{C}=\frac{\sqrt{\tilde{\omega}\left(-Z\sigma\sqrt{3}+3\tilde{\omega}\right)}}{\sqrt{3}},\quad C_p=\frac{3^{\frac{3}{8}}(\tilde{\omega}\sigma Z)^{\frac{1}{4}}\sqrt{C}}{2^{\frac{3}{4}}},\quad C_q=\frac{3^{\frac{7}{8}}(2\tilde{\omega}\sigma Z)^{\frac{1}{4}}\sqrt{C}}{6},$$

from which (8) follows. It remains to show that (p^*,q^*) is a local minimum, i.e. for any variation $(\delta p,\delta q)$ and $k\in\{\bar{k},k_{\max}\}$, we must have

$$\rho(k,\tilde{\omega},\sigma,Z,0,p^*+\delta p+i(q^*+\delta q))\geq\rho(k,\tilde{\omega},\sigma,Z,0,p^*+iq^*).$$

By the Taylor formula, it suffices to prove that there is no variation $(\delta p,\delta q)$ such that for $k\in\{\bar{k},k_{\max}\}$

$$\delta p\frac{\partial\rho}{\partial p}(k,\tilde{\omega},\sigma,Z,0,p^*+iq^*)+\delta q\frac{\partial\rho}{\partial q}(k,\tilde{\omega},\sigma,Z,0,p^*+iq^*)<0. \tag{10}$$

We prove this by contradiction, and it is necessary to obtain the next higher order terms in the expansions of p^*, q^* and $\bar{k}$. After a lengthy computation, we find that asymptotically

$$\bar{k}\sim\bar{C}+\tilde{C}h,\quad p^*\sim C_ph^{-\frac{1}{2}}+\tilde{C}_ph^{\frac{3}{2}},\quad q^*\sim C_qh^{-\frac{1}{2}}+\tilde{C}_qh^{\frac{1}{2}}.$$

The computation of these new three constants allows us to obtain the partial derivatives of ρ

$$\begin{aligned}&\tfrac{\partial\rho}{\partial p}(\bar{k})\sim\tfrac{2}{\bar{C}}h,\ \tfrac{\partial\rho}{\partial q}(\bar{k})\sim-\tfrac{3^{\frac{1}{4}}(2\omega\sigma\mu)^{\frac{1}{2}}}{C^2}h^2,\\ &\tfrac{\partial\rho}{\partial p}(k_{\max})\sim-\tfrac{2}{\bar{C}}h,\ \tfrac{\partial\rho}{\partial q}(k_{\max})\sim\tfrac{3^{\frac{1}{4}}(2\omega\sigma\mu)^{\frac{1}{2}}}{C^2}h^2.\end{aligned}$$

Introducing these results into (10), we get $\delta p\frac{2}{\bar{C}}h-\delta q\frac{3^{\frac{1}{4}}(2\omega\sigma\mu)^{\frac{1}{2}}}{C^2}h^2<0$ and -$\delta p\frac{2}{\bar{C}}h+\delta q\frac{3^{\frac{1}{4}}(2\omega\sigma\mu)^{\frac{1}{2}}}{C^2}h^2<0$, clearly a contradiction, and thus (p^*,q^*) is a local minimum.

We see that for h small, both the one parameter and two parameter transmission conditions can be written as $\rho_1^*=1-\alpha_1\sqrt{h}+O(h)$ and $\rho_2^*=1-\alpha_2\sqrt{h}+O(h)$. The ratio $\frac{\alpha_2}{\alpha_1}$ is equal to $3^{\frac{3}{8}}/\sqrt{2}\approx 1.067$, which shows that the convergence factors are almost equal. Hence the hypothesis $p=q$, used in [4] to simplify the analysis, is justified.

We treat now the overlapping case of (5), with an overlap of one mesh size.

Theorem 5. *For $\sigma > 0$ and $L = h$, a local minimum (p^*, q^*) of (5) is for h small given by*

$$p^* = \frac{3^{\frac{1}{2}}(\omega\sigma\mu)^{\frac{1}{3}}}{2^{\frac{4}{3}}h^{\frac{1}{3}}}, \quad q^* = \frac{(\omega\sigma\mu)^{\frac{1}{3}}}{2^{\frac{4}{3}}h^{\frac{1}{3}}}, \quad \rho^*_{2L} = 1 - 2^{\frac{5}{6}}3^{\frac{3}{8}}(\omega\sigma\mu)^{\frac{1}{6}}h^{\frac{1}{3}} + O(h^{\frac{2}{3}}). \quad (11)$$

Proof. As in the proof of Theorem 4, we first observe numerically that the solution of (5) equioscillates once, i.e. (p^*, q^*) is solution of

$$\rho(\bar{k}_1, \tilde{\omega}, \sigma, Z, h, p^* + iq^*) = \rho(\bar{k}_2, \tilde{\omega}, \sigma, Z, h, p^* + iq^*),$$

where $\bar{k}_1$ and $\bar{k}_2$ are interior local maxima of ρ, and we obtain asymptotically for h small

$$\bar{k}_1 \sim C_{b_1}, \bar{k}_2 \sim C_{b_2}h^{-\frac{2}{3}}, p^* \sim C_p h^{-\frac{1}{3}} \text{ and } q^* \sim C_q h^{-\frac{1}{3}}.$$

It remains to find C_{b_1}, C_{b_2}, C_p and C_q. Proceeding as before, we obtain four equations from the necessary conditions of a minimum, with solution

$$C_p = \frac{3^{\frac{1}{2}}(2\omega\sigma\mu)^{\frac{1}{2}}}{2}, C_q = \frac{C_p}{\sqrt{3}}, C_{b_1} = \frac{\sqrt{\tilde{\omega}\left(-Z\sigma\sqrt{3} + 3\tilde{\omega}\right)}}{\sqrt{3}}, C_{b_2} = \sqrt{2C_p},$$

which leads to (11). To prove that (p^*, q^*) is a local minimum, proceeding as before, we obtain after a lengthy computation the higher order expansion

$$\bar{k}_1 \sim C_{b_1} + \tilde{C}_{b_1}h^{\frac{2}{3}}, \bar{k}_2 \sim C_{b_2}h^{-\frac{2}{3}} + \tilde{C}_{b_2}, p^* \sim C_p h^{-\frac{1}{3}} + \tilde{C}_p h^{\frac{1}{3}}, q^* \sim C_q h^{-\frac{1}{3}} + \tilde{C}_q h^{\frac{1}{3}}.$$

The computation of these four new constants allows us then to obtain the partial derivatives of ρ,

$$\begin{aligned} \frac{\partial\rho}{\partial p}(\bar{k}_1) &\sim \frac{8\cdot 2^{\frac{1}{6}}h^{\frac{2}{3}}}{3^{\frac{1}{4}}(\omega\sigma\mu)^{\frac{1}{6}}}, \frac{\partial\rho}{\partial q}(\bar{k}_1) \sim -\frac{2\cdot 2^{\frac{5}{6}}(\omega\sigma\mu)^{\frac{1}{6}}h^{\frac{4}{3}}}{3^{\frac{1}{4}}}, \\ \frac{\partial\rho}{\partial p}(\bar{k}_2) &\sim -\frac{4\cdot 2^{\frac{1}{6}}h^{\frac{2}{3}}}{3^{\frac{1}{4}}(\omega\sigma\mu)^{\frac{1}{6}}}, \frac{\partial\rho}{\partial q}(\bar{k}_2) \sim \frac{2^{\frac{5}{6}}(\omega\sigma\mu)^{\frac{1}{6}}h^{\frac{4}{3}}}{3^{\frac{1}{4}}}. \end{aligned} \quad (12)$$

In order to reach a contradiction, we assume again there exists, by the Taylor theorem, a variation $(\delta p, \delta q)$ such that $\delta p\frac{\partial\rho}{\partial p}(k, \tilde{\omega}, \sigma, Z, h, p^* + iq^*) + \delta q\frac{\partial\rho}{\partial q}(k, \tilde{\omega}, \sigma, Z, h, p^* + iq^*) < 0$, for $k \in \{\bar{k}_1, k_2\}$. Using (12), we get $8\frac{2^{\frac{1}{6}}h^{\frac{2}{3}}}{3^{\frac{1}{4}}(\omega\sigma\mu)^{\frac{1}{6}}}\delta p - 2\frac{2^{\frac{5}{6}}(\omega\sigma\mu)^{\frac{1}{6}}h^{\frac{4}{3}}}{3^{\frac{1}{4}}}\delta q < 0$ and $-4\frac{2^{\frac{1}{6}}h^{\frac{2}{3}}}{3^{\frac{1}{4}}(\omega\sigma\mu)^{\frac{1}{6}}}\delta p + \frac{2^{\frac{5}{6}}(\omega\sigma\mu)^{\frac{1}{6}}h^{\frac{4}{3}}}{3^{\frac{1}{4}}}\delta q < 0$, clearly a contradiction, and thus (p^*, q^*) is a local minimum.

We also observe in this case that for h small, both convergence factors can be written as $\rho^*_{1L} = 1 - \alpha_{1L}\, h^{\frac{1}{3}} + O(h^{\frac{2}{3}})$ and $\rho^*_{2L} = 1 - \alpha_{2L}h^{\frac{1}{3}} + O(h^{\frac{2}{3}})$, and the ratio $\frac{\alpha_{2L}}{\alpha_{1L}}$ is equal to $3^{\frac{1}{4}}/2^{\frac{1}{3}} \approx 1.044$, hence both convergence factors are almost equal. We show an example of these convergence factors in Fig. 1.

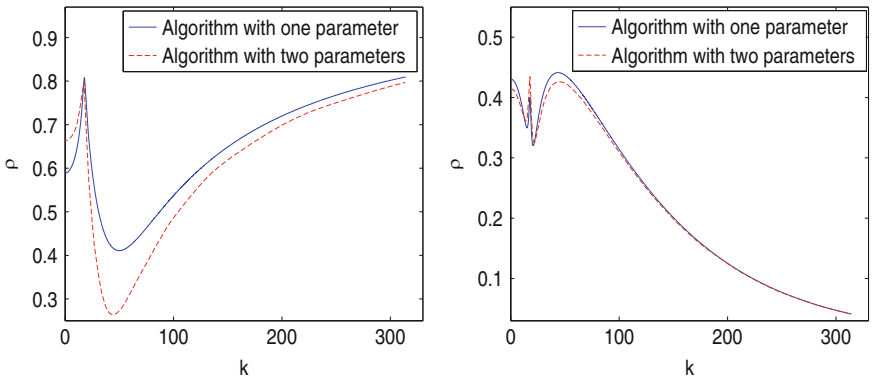


Fig. 1. Convergence factor comparison of algorithms with one and two parameters for $\omega = 2\pi$, $\sigma = 2$ and $\mu = \varepsilon = 1$, for the non-overlapping case, $L = 0$, on the *left*, and the overlapping case, $L = h = \frac{1}{100}$, on the *right*

4 Numerical Results

We present now a numerical test in order to compare the performance of both the one and two parameter algorithms. We compute the propagation of a plane wave in a heterogeneous medium. The domain is $\Omega = (-1,1)^2$. The relative permittivity and the conductivity of the background media is $\varepsilon_1 = 1.0$ and $\sigma_1 = 1.8$, while that of the square material inclusion is $\varepsilon_2 = 8.0$ and $\sigma_2 = 7.5$, see the left picture of Fig. 2. The magnetic permeability μ is constant in Ω and we impose on the boundary an incident field $(H_x^{inc}, H_y^{inc}, E_z^{inc})$. The domain Ω is decomposed into two subdomains $\Omega_1 = (-1, L) \times (-1, 1)$ and $\Omega_2 = (0, 1) \times (-1, 1)$; L is the overlapping size and is equal to the mesh size. We use, in each subdomain, a discontinuous Galerkin method (DG) with a uniform polynomial approximation of order one, two and three, denoted by *DG-P*1, *DG-P*2 and *DG-P*3, see [5]. The results are shown in Fig. 3, and are in good agreement with our analytical results.

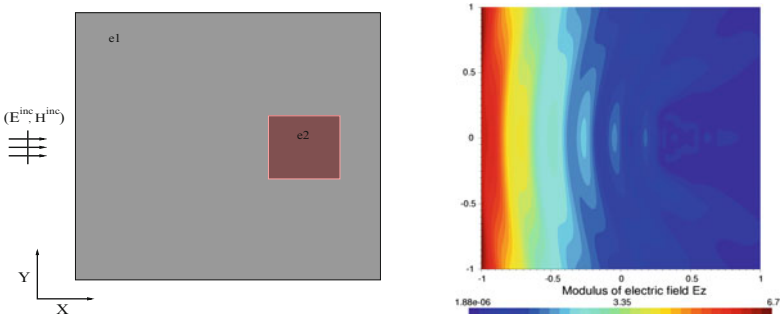


Fig. 2. Configuration of our test problem on the *left*, and the numerical solution on the *right*

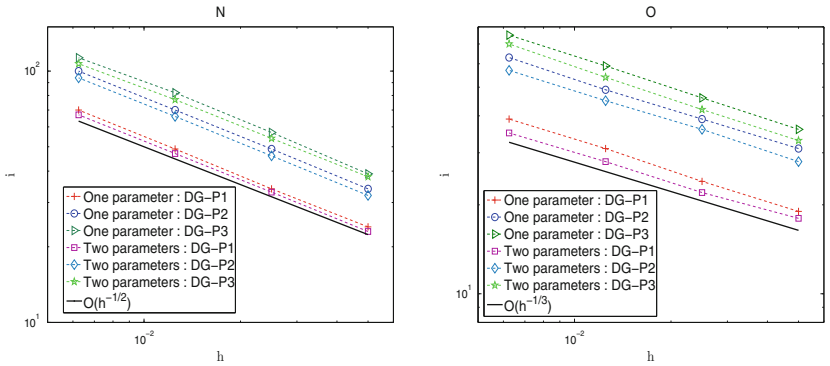


Fig. 3. Number of iterations against the mesh size h, to attain a relative residual reduction of 10^{-8}

5 Conclusion

We compared in this paper a one and a two parameter family of transmission conditions for optimized Schwarz methods applied to Maxwell's equations. Our asymptotic analysis reveals that the addition of a second parameter does not lead to a significant improvement of the algorithm, and it is therefore justified to consider only the simpler case of a one parameter family of transmission conditions. These results are also confirmed by our numerical experiments.

Bibliography

[1] A. Alonso-Rodriguez and L. Gerardo-Giorda. New nonoverlapping domain decomposition methods for the harmonic Maxwell system. *SIAM J. Sci. Comput.*, 28(1):102–122, 2006.

[2] P. Chevalier and F. Nataf. An OO2 (Optimized Order 2) method for the Helmholtz and Maxwell equations. In *10th International Conference on Domain Decomposition Methods in Science and in Engineering*, pages 400–407, Boulder, Colorado, USA, 1997. AMS.

[3] P. Collino, G. Delbue, P. Joly, and A. Piacentini. A new interface condition in the non-overlapping domain decomposition for the Maxwell equations. *Comput. Methods Appl. Mech. Engrg.*, 148:195–207, 1997.

[4] V. Dolean, M. El Bouajaji, M. Gander, and S. Lanteri. Optimized Schwarz methods for Maxwell's equations with non-zero electric conductivity. In *Domain Decomposition Methods in Science and Engineering XIX*, LNCSE, Vol. 78, pp 269–276, 2011.

[5] V. Dolean, M. El Bouajaji, M. Gander, S. Lanteri, and R. Perrussel. Domain decomposition methods for electromagnetic wave propagation problems in heterogeneous media and complex domains. In *Domain Decomposition Methods in Science and Engineering XIX*,LNCSE, Vol. 78, pp 15–26, 2011.

[6] V. Dolean, L. Gerardo-Giorda, and M. J. Gander. Optimized Schwarz methods for Maxwell equations. *SIAM J. Scient. Comp.*, 31(3):2193–2213, 2009.

[7] V. Dolean, S. Lanteri, and R. Perrussel. Optimized Schwarz algorithms for solving time-harmonic Maxwell's equations discretized by a discontinuous Galerkin method. *IEEE. Trans. Magn.*, 44(6):954–957, 2008.

[8] V. Dolean, M. El Bouajaji, M. J. Gander, S. Lanteri, Optimized Schwarz methods for the time-harmonic Maxwell equations with damping, *SIAM J. Sci. Comput.*, Vol 34., No. 4, pp. 2048–2071, 2012.

[9] Zhen Peng and Jin-Fa Lee. Non-conformal domain decomposition method with second-order transmission conditions for time-harmonic electromagnetics. *J. Comput. Phys.*, 229(16):5615–5629, 2010.

Hybrid Domain Decomposition Solvers for the Helmholtz and the Time Harmonic Maxwell's Equation

M. Huber[1], A. Pechstein[2] and J. Schöberl[1]

[1] Institute for Analysis and Scientific Computing, Wiedner Hauptstrasse 8-10, A-1040 Wien
martin.huber@tuwien.ac.at, joachim.schoeberl@tuwien.ac.at
[2] Institute for Technical Mechanics, Altenbergerstrasse 69, A-4040 Linz
astrid.pechstein@jku.at

Summary. We present hybrid finite element methods for the Helmholtz equation and the time harmonic Maxwell equations, which allow us to reduce the unknowns to degrees of freedom supported only on the element facets and to use efficient iterative solvers for the resulting system of equations. For solving this system, additive and multiplicative Schwarz preconditioners with local smoothers and a domain decomposition preconditioner with an exact subdomain solver are presented. Good convergence properties of these preconditioners are shown by numerical experiments.

1 Introduction

When solving the Helmholtz equation with a standard finite element method (FEM), due to the oscillatory behaviour of the solution and the pollution error [8] a large number of degrees of freedom (DoFs) is needed to resolve the wave, especially for high wave numbers. To overcome this difficulty, many methods have been developed during the last years. Apart from *hp* FEM [8], Galerkin Least Square Methods [7] or Discontinuous Galerkin Methods [6], some methods make use of problem adapted functions like plane waves. The most popular among them are the Partition of Unity Method [9], the Discontinuous Enrichment Approach [5] or the UWVF [2, 10]. All these techniques end up with large, complex valued, indefinite, possible symmetric linear systems. Although some advances have been made [3, 4], efficient preconditioners for wave type problems are still a big challenge.

In the present work the hybrid FEM from [11] is used for the Helmholtz equation and extended to the Maxwell case. This method allows us to use efficient iterative methods for solving the resulting linear system of equations. Following hybridization techniques from [1], the tangential continuity of the flux field is broken across element interfaces. In order to impose continuity again, Lagrange multipliers supported only on the facets, which can be interpreted as the tangential component of the unknown field, are introduced. Adding a second set of Lagrange multipliers,

R. Bank et al. (eds.), *Domain Decomposition Methods in Science and Engineering XX*, Lecture Notes in Computational Science and Engineering 91,
DOI 10.1007/978-3-642-35275-1_32,

representing the tangential component of the flux field, allows us, due to local Robin boundary conditions, to eliminate the volume DoFs. Because, after hybridization, there is no coupling between volume basis functions of different elements, elimination of the volume DoFs can be done cheaply element by element, and the system of equation is reduced onto the smaller set of Lagrange multipliers. For the reduced system we present additive (AS) and multiplicative Schwarz (MS) block preconditioners with blocks related to DoFs of one facet and element, respectively. Additionally a domain decomposition (DD) preconditioner, which directly solves for the DoFs belonging to one subdomain, is investigated. This preconditioner is especially advantageous for domains contains cavity like structures. Numerical tests show, that a preconditioned CG iteration has good convergence properties combined with these preconditioners.

2 Hybridization of the Wave Equations

In the sequence, we will stick to the following settings. As computational domain we consider a Lipschitz polyhedron $\Omega \subset \mathbb{R}^d$ with $d = 2,3$ and the boundary $\Gamma = \partial\Omega$. In the scalar case, we search for a function $u : \Omega \to \mathbb{C}$ and a vector valued field $\mathbf{v} : \Omega \to \mathbb{C}^d$, which fulfills the Helmholtz equation in mixed form

$$\operatorname{grad} u = i\omega\mathbf{v} \qquad \text{and} \qquad \operatorname{div}\mathbf{v} = i\omega u \qquad \text{in } \Omega$$

with absorbing boundary conditions $\mathbf{v}\cdot\mathbf{n} + u = g$ on Γ, where ω is the angular frequency and $\mathbf{n}$ the outer normal vector. From [9] we know, that the solution u exists and is unique.

In the vectorial case, i.e. the harmonic Maxwell's equations, we search for a vector valued function $\mathbf{E} : \Omega \to \mathbb{C}^3$ and a flux field $\mathbf{H} : \Omega \to \mathbb{C}^3$, which solves

$$\operatorname{curl}\mathbf{H} + i\omega\mathbf{E} = 0 \qquad \text{and} \qquad \operatorname{curl}\mathbf{E} - i\omega\mathbf{H} = 0 \qquad \text{in } \Omega$$

under the boundary condition $-\mathbf{n}\times\mathbf{H} + \mathbf{E}_{\|} = \mathbf{g}$ on Γ, where $\mathbf{E}_{\|}$ represents the tangential component of $\mathbf{E}$, i.e. $\mathbf{n}\times\mathbf{E}\times\mathbf{n}$.
When deriving the hybrid formulation, we use a regular finite element mesh $\mathscr{T}$ with elements T, and the set of facets is called $\mathscr{F}$. The vector $\mathbf{n}_T$ is the outer normal vector of the element T, and $\mathbf{n}_F$ represents the normal vector onto a facet F. Furthermore, we denote a volume integral as $(u,v)_T := \int_T uv \,\mathrm{d}\mathbf{x}$, and a surface integral as $\langle u,v\rangle_{\partial T} := \int_{\partial T} uv \,\mathrm{d}s$.

2.1 The Mixed Hybrid Formulation for the Helmholtz Equation

The mixed hybrid formulation for the Helmholtz equation was already introduced in [11]. For completeness, we repeat the problem formulation:
Find $(u,\mathbf{v},u^F,v^F) \in L^2(\Omega)\times H(\operatorname{div},T)\times L^2(\mathscr{F})\times L^2(\mathscr{F}) =: X\times\tilde{Y}\times X^F\times Y^F$, such that for all $(\sigma,\mathbf{w},\sigma^F,w^F) \in X\times\tilde{Y}\times X^F\times Y^F$

$$\sum_{T\in\mathscr{T}}\Big(\big(i\omega u,\sigma\big)_T-\big(i\omega\mathbf{v},\mathbf{w}\big)_T-\big(\operatorname{div}\mathbf{v},\sigma\big)_T-\big(u,\operatorname{div}\mathbf{w}\big)_T+\big\langle u^F,\mathbf{n}_T\cdot\mathbf{w}\big\rangle_{\partial T}$$
$$+\big\langle\mathbf{n}_T\cdot\mathbf{v},\sigma^F\big\rangle_{\partial T}+\big\langle\mathbf{n}_F\cdot\mathbf{v}-v^F,\mathbf{n}_F\cdot\mathbf{w}-w^F\big\rangle_{\partial T}\Big)+\big\langle u^F,\sigma^F\big\rangle_\Gamma=\big\langle g,\sigma^F\big\rangle_\Gamma.$$

2.2 The Mixed Hybrid Formulation for the Maxwell Problem

We will now concentrate on the derivation of the mixed hybrid formulation for the vectorial wave equation. We start from the mixed system of equations from above, multiply the first equation with a test function $\mathbf{e}\in U:=(L^2(\Omega))^3$ and the second one with a function $\mathbf{h}\in V:=H(\operatorname{curl},\Omega)$ and integrate over the domain Ω. Performing integration by parts elementwise leads to

$$\sum_{T\in\mathscr{T}}\Big(\big(\operatorname{curl}\mathbf{H},\mathbf{e}\big)_T+\big(i\omega\mathbf{E},\mathbf{e}\big)_T\Big)=0\qquad\forall\mathbf{e}\in U$$
$$\sum_{T\in\mathscr{T}}\Big(\big(\mathbf{E},\operatorname{curl}\mathbf{h}\big)_T-\big(i\omega\mathbf{H},\mathbf{h}\big)_T-\big\langle\mathbf{E},\mathbf{n}_T\times\mathbf{h}\big\rangle_{\partial T}\Big)=0\qquad\forall\mathbf{h}\in V.$$

Note that for a tangential continuous field $\mathbf{E}$, i.e. $\mathbf{n}\times\mathbf{E}\times\mathbf{n}$ is continuous on element interfaces, the boundary integrals for inner facets cancel due to the tangential continuity of $\mathbf{h}$, and inserting the absorbing boundary condition into the boundary facet integrals leads to the standard mixed finite element formulation for our problem.

Next, the tangential continuity of the flux field $\mathbf{H}$ is broken across element interfaces, thus we search for $\mathbf{H}\in\tilde{V}:=\big\{\mathbf{v}\in(L^2(\Omega))^3\ :\ \mathbf{v}|_T\in H(\operatorname{curl},T)\ \forall T\in\mathscr{T}\big\}$. In order to reinforce continuity, Lagrange multipliers $\mathbf{E}^F$, which are only supported on the element facets, i.e. they are from the space $U^F:=(L^2(\mathscr{F}))^3$, are introduced. The continuity of the tangential fluxes is reached via an additional equation, which forces the jump of $[\mathbf{n}\times\mathbf{H}]:=\mathbf{n}_{T_1}\times\mathbf{H}|_{T_1}+\mathbf{n}_{T_2}\times\mathbf{H}|_{T_2}$ for inner facets $F\in\mathscr{F}_I$ with adjacent elements T_1 and T_2 to zero, thus

$$\sum_{F\in\mathscr{F}_I}\big\langle[\mathbf{n}\times\mathbf{H}],\mathbf{e}\big\rangle_F=\sum_{T\in\mathscr{T}}\Big(\big\langle\mathbf{n}_T\times\mathbf{H},\mathbf{e}\big\rangle_{\partial T}-\big\langle\mathbf{n}_T\times\mathbf{H},\mathbf{e}\big\rangle_{\partial T\cap\Gamma}\Big)=0,\quad\forall\mathbf{e}\in U^F.$$

The resulting system of equations for $(\mathbf{E},\mathbf{H},\mathbf{E}^F)\in U\times\tilde{V}\times U^F$ reads as

$$\sum_{T\in\mathscr{T}}\Big(\big(\operatorname{curl}\mathbf{H},\mathbf{e}\big)_T+\big(i\omega\mathbf{E},\mathbf{e}\big)_T\Big)=0\qquad\forall\mathbf{e}\in U$$
$$\sum_{T\in\mathscr{T}}\Big(\big(\mathbf{E},\operatorname{curl}\mathbf{h}\big)_T-\big(i\omega\mathbf{H},\mathbf{h}\big)_T-\big\langle\mathbf{E}^F,\mathbf{n}_T\times\mathbf{h}\big\rangle_{\partial T}\Big)=0\qquad\forall\mathbf{h}\in\tilde{V}$$
$$-\sum_{T\in\mathscr{T}}\big\langle\mathbf{n}_T\times\mathbf{H},\mathbf{e}^F\big\rangle_{\partial T}+\big\langle\mathbf{E}^F,\mathbf{e}^F\big\rangle_\Gamma=\big\langle\mathbf{g},\mathbf{e}^F\big\rangle_\Gamma\qquad\forall\mathbf{e}^F\in U^F.$$

In this system of equations, the Lagrange parameter $\mathbf{E}^F$ plays the role of the tangential component of $\mathbf{E}$, evaluated on the facets. Because there is no coupling between volume DoFs belonging to different elements, it is possible to eliminate the volume unknowns $\mathbf{E}$ and $\mathbf{H}$, cheaply by static condensation (compare [1]). The resulting system of equations needs now to be solved only for the Lagrange multipliers.

In order to eliminate the inner DoFs, one has to solve the first two equations of the system from above for some function $\mathbf{E}^F$ element by element. But this is equivalent to solving a Dirichlet problem, and uniqueness of the solution can not be guaranteed. This drawback can be compensated by adding a new facet unknown $\mathbf{H}^F \in V^F := (L^2(\mathscr{F}))^3$ representing $\mathbf{n}_F \times \mathbf{H}$ on the facets via a consistent stabilization term $\sum_T \langle \mathbf{n}_F \times \mathbf{H} - \mathbf{H}^F, \mathbf{n}_F \times \mathbf{h} - \mathbf{h}^F \rangle_{\partial T}$. We obtain

$$\sum_{T \in \mathscr{T}} \Big(\big(\operatorname{curl} \mathbf{H}, \mathbf{e}\big)_T + \big(i\omega \mathbf{E}, \mathbf{e}\big)_T \Big) = 0 \qquad \forall \mathbf{e} \in U \tag{1}$$

$$\sum_{T \in \mathscr{T}} \Big(\big(\mathbf{E}, \operatorname{curl} \mathbf{h}\big)_T - \big(i\omega \mathbf{H}, \mathbf{h}\big)_T - \langle \mathbf{E}^F, \mathbf{n}_T \times \mathbf{h} \rangle_{\partial T} - \langle \mathbf{n}_T \times \mathbf{H}, \mathbf{n}_T \times \mathbf{h} \rangle_{\partial T} + \langle \mathbf{H}^F, \mathbf{n}_F \times \mathbf{h} \rangle_{\partial T} \Big) = 0 \qquad \forall \mathbf{h} \in \tilde{V} \tag{2}$$

$$\sum_{T \in \mathscr{T}} \Big(\langle \mathbf{n}_F \times \mathbf{H}, \mathbf{h}^F \rangle_{\partial T} - \langle \mathbf{H}^F, \mathbf{h}^F \rangle_{\partial T} \Big) = 0 \qquad \forall \mathbf{h}^F \in V^F \tag{3}$$

$$- \sum_{T \in \mathscr{T}} \langle \mathbf{n}_T \times \mathbf{H}, \mathbf{e}^F \rangle_{\partial T} + \langle \mathbf{E}^F, \mathbf{e}^F \rangle_\Gamma = \langle \mathbf{g}, \mathbf{e}^F \rangle_\Gamma \qquad \forall \mathbf{e}^F \in U^F. \tag{4}$$

Now, by static condensation the time harmonic Maxwell's equation with absorbing boundary conditions has to be solved on the element level, where uniqueness is guaranteed, and the resulting system contains only the facet unknowns $\mathbf{E}^F$ and $\mathbf{H}^F$. Thus we search for a function $\mathbf{w} \in W := U^F \times V^F$ such that

$$s(\mathbf{w}, \mathbf{v}) = f(\mathbf{v}) \qquad \forall \mathbf{v} \in W,$$

where the Schur complement bilinearform s and the linearform f are obtained from (1) to (4) by eliminating the unknowns $\mathbf{E}$ and $\mathbf{H}$. Elimination of the inner DoFs can be also seen as calculating for a given incoming impedance trace $\mathbf{E}^F - \mathbf{H}^F$ the resulting outgoing impedance trace $\mathbf{E}^F + \mathbf{H}^F$ on the element level. By exchanging the Dirichlet and Neumann traces $\mathbf{E}^F, \mathbf{H}^F$ by incoming and outgoing impedance traces, one obtains an equivalent formulation which fits well into the context of the UWVF of [2].

3 Iterative Solvers

In this section, we focus on solving the system of equations. As already mentioned, the volume DoFs can be eliminated cheaply element by element, and the resulting system of equation just has to be solved for the much smaller number of facet DoFs. Because volume DoFs of one element couple apart from themselves only to facet DoFs of the surrounding facets, the Schur complement matrix S obtained by static condensation is sparse, and it just has nonzero entries between facet DoFs belonging to facets of the same element. Due to the hybrid formulation, efficient iterative solvers can be used for the reduced system of equations.

Because the Schur complement matrix is complex symmetric, a preconditioned CG-iteration together with an AS or MS block preconditioner, M_{AS} and M_{MS} is used,

although convergence for complex symmetric matrices is not guaranteed. The iteration matrices of these two preconditioners are given as

$$I - M_{AS}^{-1} S = I - \sum_{i=1}^{n} P_i,$$

$$I - M_{MS}^{-1} S = \Big(\prod_{i=n}^{1} (I - P_i) \Big) \Big(\prod_{i=1}^{n} (I - P_i) \Big),$$

where P_i is the matrix representation of the variational projector $\mathscr{P}_i : W \to W_i \subset W$ with respect to the bilinearform s. In the scalar case $W = X^F \times Y^F$. We will use two different choices of subspaces W_i, functions supported on the facet F_i or on facets, which are boundary facets of the element T_i. Note that the first strategy leads to nonoverlapping blocks, while the blocks of the second choice overlap.

Apart from an AS or MS Preconditioner, a DD preconditioner compareable to [12] was used, which is based on a partitioning of the domain Ω into N subdomains Ω_i. The iteration matrix of this preconditioner can be described by

$$I - M_{DD}^{-1} S = \Big(\prod_{i=n}^{1} (I - P_{I,i}) \Big) \Big(I - \sum_{i=1}^{N} P_{\Omega_i} \Big) \Big(\prod_{i=1}^{n} (I - P_{I,i}) \Big),$$

where P_{Ω_i} and $P_{I,i}$ are matrices corresponding to variational projection operators which project to the spaces W_{Ω_i} and $W_{I,i}$. The space W_{Ω_i} contains functions which are supported only on facets in the interior of the subdomain Ω_i, while the space $W_{I,i}$ is choosen such that it contains functions which are only supported on facets of an element T_i such that $\partial T_i \cap \partial \Omega_j \neq \emptyset$. Again a nonoverlapping option is to collect the functions supported on a facet F_i which is located on Γ or the subdomain interfaces in $W_{I,i}$. Thus, in each preconditioner step a forward block Gauss Seidel iteration is carried out, followed by a direct inversion of each subdomain block and a backward block Gauss Seidel step. Note that solving directly for the unknowns in a subdomain is equivalent to solve a problem with robin boundary conditions on the subdomain, and uniqueness and existence are guaranteed.

One big advantage of the DD preconditioner is, that it can cope with problems containing cavity like structures. For such problems other preconditioners suffer from internal reflections, which leads to high iteration numbers. If the whole cavity is contained in one single subdomain Ω_i , the DD preconditioner inverts the whole matrix block related to the cavity, and internal reflections are treated exactly. Thus they do not influence the iteration number.

4 Numerical Results

In order to demonstrate the dependence of the number of iterations on polynomial order, wavelength and meshsize h for the presented preconditioners, we choose a simple two dimensional model problem with a wave of Gaussian amplitude and wavelength λ propagating through a unit square domain (compare Fig. 1). For a

meshsize $h = \lambda = 0.1$ the lefthand plot shows the number of iterations for different polynomial orders. For the three preconditioners, the DoFs of an element were collected in one block. In addition, for the DD preconditioner, the computational domain was divided into nine subdomains. If the polynomial order is large enough to resolve the wave, i.e. larger than four, the number of iterations stays constant or is only slightly growing with growing polynomial order, while the number of facet unknowns grows linearly in 2D.

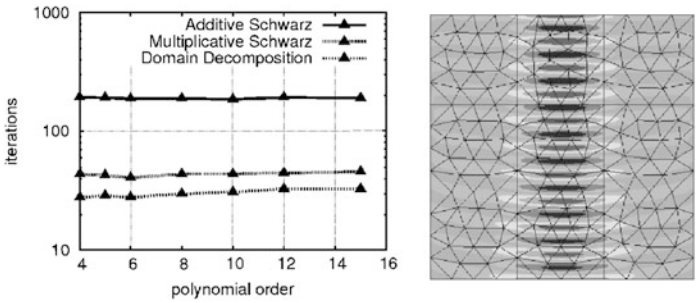


Fig. 1. Iterations depending on the polynomial order (*left*) for the 2D model problem (*right*)

Table 1. Iterations depending on wavelength and mesh size for the MS/DD Preconditioner ($p = 6$).

λ	0.64	0.32	0.16	0.08	0.04	0.02	0.01
$h = 0.16$	35/40	35/38	32/33	31/31			
$h = 0.08$	52/42	48/38	50/36	47/33	50/38		
$h = 0.04$	88/55	76/47	74/43	76/39	65/35	97/59	
$h = 0.02$	147/75	129/55	113/48	117/44	118/42	115/38	199/82
$h = 0.01$	246/107	236/80	226/60	203/53	228/49	271/50	291/45

Next we investigate the dependence on h and λ for a fixed polynomial order of 6. The results are presented in Table 1. For λ smaller than $\frac{h}{2}$, which corresponds to less than three unknowns per wavelength, the solution can not be resolved, and the solvers show large iteration numbers. Fixing h, the iteration number is minimal at about $h \approx \lambda$, i.e. at about six unknowns per wavelength, and it increases for growing wavelength. For $h = 0.16$ every subdomain consists of only a small number of elements, and an inversion of the DoFs subdomain by subdomain is compareable to an inversion element by element. Therefore the two preconditioners show about the same performance. If h decreases, it is more and more advantageous to collect the unknowns in subdomain blocks. While the iteration number almost doubles for the MS preconditioner if the mesh size is divided by 2, the increase is much

less for the DD preconditioner. Table 2 shows, that the DD preconditioner also performs better than the MS preconditioner with respect to time, although one iteration is more expensive.

Table 2. Iteration times for $\lambda = 0.08$ and a polynomial order of 6.

h	DoFs	MS	DD
0.16	69980	0.35	0.37
0.08	217900	1.73	1.33
0.04	701228	9.30	5.15
0.02	2518524	53.5	22.4
0.01	9857920	367	111

Table 3. Iteration numbers and computational times for the cavity and the square.

	cavity		square	
	its.	time(s)	its.	time(s)
DD (element)	35	40.4	34	31.2
DD (facet)	64	69.7	61	59.7
MS (element)	1612	1720	102	88.9
AS (element)	$> 10^5$	$>$ 1h	575	186

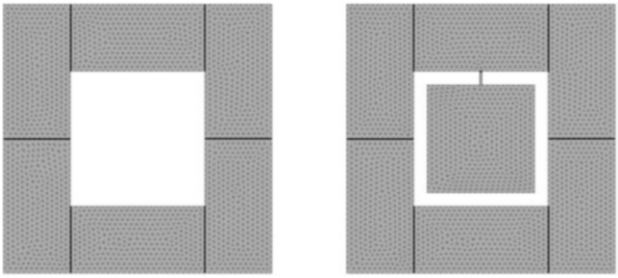


Fig. 2. A resonator (*right*) is compared with the domain without cavity (*left*)

Now we compare the preconditioners for a resonator and the domain without cavity (compare Fig. 2). From the top of the square an incident wave with $\lambda = 0.01$ is prescribed. The DD-preconditioner uses, depending on the presence of the cavity six and seven subdomains, respectively, where all cavity DoFs, including the cavity boundary are collected in one single block. Table 3 shows the iteration numbers and computational times for different preconditioners and for the two examples. For the domain without cavity the performance of the preconditioners is compareable. When the cavity is added, reflections inside the cavity lead to an enormous increase in iteration numbers and computational times for the AS and the MS preconditioner. Because of direct inversion of the cavity DoFs, the DD preconditioner does not suffer from internal reflections and the iteration number stays almost constant, which leads together with a larger number of unknowns to a moderate increase in computational time.

We finish the numerical results section with an example from optics. A small sphere with radius 0.3 and refractive index 2 is placed (not exactly in the center) in a spherical computational domain with radius 1 and background refractive index 1.

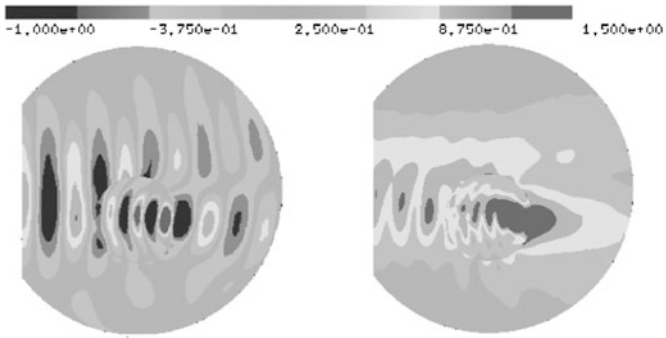


Fig. 3. Real part of E_y (*left*) and $|E|$ (*right*) evaluated at a cross section parallel to the xy plane

We prescribe an incident wave from the left with a Gaussian amplitude and wavelength 0.35, such that the diameter of the computational domain is approximately six wavelength in free space. In order to resolve the wave we used 3,256 elements with a polynomial order of 6, which results in 1.66 millions of unknowns. The solution (compare Fig. 3) was obtained by 258 cg-iterations with a Block AS preconditioner.

Bibliography

[1] D.N. Arnold and F. Brezzi. Mixed and nonconforming finite element methods: Implementation, postprocessing and error estimates. *RAIRO Model. Math. Anal. Numer.*, 19(1):7–32, 1985.

[2] O. Cessenat and B. Despres. Application of an Ultra Weak Variational Formulation of elliptic PDEs to the two-dimensional Helmholtz problem. *SIAM J. Numer. Anal.*, 35(1):255–299, 1998.

[3] B. Engquist and L. Ying. Sweeping preconditioner for the Helmholtz equation: Hierarchical matrix representation. *Comm. Pure Appl. Math.*, 64(5):697–735, 2011.

[4] Y.A. Erlangga, C. Vuik, and C.W. Oosterlee. On a class of preconditioners for solving the Helmholtz equation. *Appl. Numer. Math.*, 50(3-4):409–425, 2004.

[5] C. Farhat, I. Harari, and U. Hetmaniuk. A discontinuous Galerkin method with Lagrange multipliers for the solution of Helmholtz problems in the mid-frequency regime. *Comput. Methods Appl. Mech. Engrg.*, 192(11-12):1389–1419, 2003.

[6] X. Feng and H. Wu. Discontinuous Galerkin methods for the Helmholtz equation with large wave number. *SIAM J. Numer. Anal.*, 47(4):2872–2896, 2009.

[7] I. Harari. A survey of finite element methods for time harmonic acoustics. *Comput. Methods Appl. Mech. Engrg.*, 195(13-16):1594–1607, 1997.

[8] F. Ihlenburg and I. Babuska. Finite element solution of the Helmholtz equation with high wave number part ii: *hp*-version of the FEM. *SIAM J. Numer. Anal.*, 34(1):315–358, 1997.

[9] J.M. Melenk. *On Generalized Finite Element Methods*. Phd thesis, University of Maryland, 1995.

[10] P. Monk. *Finite Element Methods for Maxwell's Equations*. Oxford University Press, Oxford, 2003.

[11] P. Monk, A. Sinwel, and J. Schöberl. Hybridizing Raviart-Thomas elements for the Helmholtz equation. *Electromagnetics*, 30(1):149–176, 2010.

[12] K. Zhao, V. Rawat, S.C. Lee, and J.F. Lee. A domain decomposition method with non-conformal meshes of finite periodic and semi-periodic structures. *IEEE Trans. Antennas and Propagation*, 55(9):2559–2570, 2007.

Multiscale Domain Decomposition Preconditioners for Anisotropic High-Contrast Problems

Yalchin Efendiev[1], Juan Galvis[1], Raytcho Lazarov[1], Svetozar Margenov[2], and Jun Ren[1]

[1] Departamento de Matemáticas, Universidad Nacional De Colombia, Colombia. jcgalvisa@unal.edu.co

[2] Acad. G. Bonchev Str., bl. 25A, Institute of Information and Communication Technologies, Bulgarian Academy of Sciences, Sofia 1113, BULGARIA.

1 Summary

In this paper, we study robust two-level domain decomposition preconditioners for highly anisotropic multiscale problems. We present a construction of coarse spaces that emploies initial multiscale basis functions and discuss techniques to achieve smaller dimensional coarse spaces without sacrificing the robustness of the preconditioner. We also present numerical results and consider possible extensions of these approaches where the dimension of the coarse space can be reduced further.

2 Introduction

Anisotropy in the diffusion arises in many applications in geosciences and engineering. In flows porous media, high anisotropy can be due to the presence of fractures that may have preferred high-conductivity directions. Because of high variations among the matrix and fracture conductivities, the permeability can have high anisotropy at the fine-scale. This is the case when fracture network conducts only in some preferred directions (e.g., in one direction in 2D problems and one or two directions in 3D problems). This preferred direction is the direction of high anisotropy and it can have heterogeneous spatial variations. For example, the presence of fracture pockets can create highly anisotropic isolated regions, while fracture corridors can form long highly anisotropic channels that span a rich hierarchy of scales. It is a challenging task to design robust preconditioners for such problems (e.g., [4]) or to solve them on a coarse grid (e.g., [2]).

In this paper, we discuss robust preconditioners for highly anisotropic multiscale diffusion problems. We assume that the high-anisotropy is also highly heterogeneous over the problem domain and these spatial variations cannot be captured within a coarse block. In the paper, robust two-level domain decomposition preconditioners

R. Bank et al. (eds.), *Domain Decomposition Methods in Science and Engineering XX*, Lecture Notes in Computational Science and Engineering 91,
DOI 10.1007/978-3-642-35275-1_33,

are constructed by designing coarse spaces that contain essential features of the fine-scale solution. The construction of the coarse spaces is based on recently introduced methods [1, 3]. We show that, for anisotropic problems, the coarse spaces can have a large dimension because fine-scale features within high-anisotropy regions need to be represented on a coarse grid. In this paper, we propose a number of remedies for this problem. Note that the proposed methods differ from existing methods for anisotropic problems [4].

The coarse spaces used in two-level domain decomposition preconditioners are constructed based on local spectral problems with a pre-computed scalar weight function. The computation of the weight function uses an initial coarse space where one basis function per coarse node is defined. We show that the local eigenvalue problem can contain many small eigenvalues, which are asymptotically vanishing as the contrast increases. One needs to include all eigenvectors that correspond to these small, asymptotically vanishing, eigenvalues. Because the number of these small eigenvalues defines the dimension of the coarse space, it is important to choose a weight function such that the dimension of the coarse space is as small as possible. If we consider the initial space as the span of piecewise (bi)linear functions, then the dimension of the coarse space can be very large. In particular, the coarse space contains all fine-scale functions with respect to the slow variable (defined as the variable representing the direction of slow conductivity) within high-anisotropy regions. On the other hand, using multiscale basis functions [2] in the initial space allows capturing the effects of high-conductivity inclusions (cf. [1, 3]) that are isolated within coarse grid blocks. As a result, the coarse space contains all fine-scale functions with respect to slow variables within high-anisotropy channels. This can lead to a substantial dimension reduction; however, unlike to the isotropic high-conductivity case, the dimension of the coarse space can still be very large as discussed in the paper. Numerical results are presented. We also discuss techniques that allow us to use smaller dimensional coarse spaces at the expenses of solving several lower dimensional problems in the channels of high-anisotropy.

3 Problem Setting and Domain Decomposition Framework

Let $D \subset \mathbb{R}^2$ (or $\mathbb{R}^3$) be a polygonal domain which is the union of a disjoint polygonal subregions $\{D_i\}_{i=1}^N$. We seek $u \in H_0^1(D)$

$$a(u,v) := \int_D \kappa(x)\nabla u \cdot \nabla v dx = \int_D f v dx, \text{ where } \kappa(x) = \begin{pmatrix} \eta(x) & 0 \\ 0 & 1 \end{pmatrix}. \qquad (1)$$

Here $\eta(x)$ is a heterogeneous field with high contrast, $\eta(x) \geq 1$. More general cases where the direction of anisotropy can change continuously in space will be considered elsewhere. Next, we introduce some notations following [1].

We assume that $\{D_i\}_{i=1}^N$ form a quasiuniform triangulation of D and denote $H = \max_i \text{diam}(D_i)$. Let $\mathscr{T}^h$ be a fine triangulation which refine $\{D_i\}_{i=1}^N$. We denote by $V^h(D)$ the usual finite element discretization of piecewise linear continuous functions with respect to the fine triangulation $\mathscr{T}^h$. Denote also by $V_0^h(D)$ the subset of $V^h(D)$ with vanishing values on ∂D. Similar notations, $V^h(\Omega)$ and $V_0^h(\Omega)$, are used for subdomains $\Omega \subset D$.

The Galerkin finite element approximation of (1) is to find $u \in V_0^h(D)$ with $a(u,v) = \int_D fv$ for all $v \in V_0^h(D)$, or in matrix form

$$Au = b, \tag{2}$$

where for all $u, v \in V^h(D)$ (considered as vectors) we have $v^T Au = a(u,v)$ and $v^T b = \int_D fv$. We assume that κ is piecewise constant coefficient in $\mathscr{T}^h$ with value $\kappa = \kappa_e = (\eta_e, 0; 0, 1)$ on each fine triangulation element $e \in \mathscr{T}^h$.

We denote by $\{D_i'\}_{i=1}^N$ the overlapping decomposition obtained from the original nonoverlapping decomposition $\{D_i\}_{i=1}^N$ by enlarging each subdomain D_i to $D_i' = D_i \cup \{x \in D, \mathrm{dist}(x, D_i) < \delta_i\}, \quad i = 1, \ldots, N$, where dist is some distance function and let $\delta = \max_{1 \le i \le N} \delta_i$. Let $V_0^h(D_i')$ be the set of finite element functions with support in D_i'. We also denote by $R_i^T : V_0^h(D_i') \to V^h(D)$ the extension by zero operator.

We use a partition of unity $\{\xi_i\}_{i=1}^N$ subordinated to the covering $\{D_i'\}_{i=1}^N$ such that

$$\sum_{i=1}^N \xi_i = 1, \quad \xi_i \in V^h(D), \quad 0 \le \xi_i \le 1 \quad \text{and} \quad \mathrm{Supp}(\xi_i) \subset D_i', \ \ i = 1, \ldots, N, \tag{3}$$

where $\mathrm{Supp}(\xi_i)$ stands for the support of the function ξ_i. This partition of unity is used to truncate global functions to local conforming functions, an essential property in the construction of a stable splitting of the space.

Given a coarse triangulation $\mathscr{T}^H$, we introduce N_c coarse basis functions $\{\Phi_i\}_{i=1}^{N_c}$. We define the coarse space by $V_0^H = \mathrm{span}\{\Phi_i\}_{i=1}^{N_c}$, and the coarse matrix $A_0 = R_0 A R_0^T$ where $R_0^T = [\Phi_1, \ldots, \Phi_{N_c}]$. We use a two level additive preconditioner of the form

$$B^{-1} = R_0^T A_0^{-1} R_0 + \sum_{i=1}^N R_i^T A_i^{-1} R_i = R_0^T A_0^{-1} R_0 + B_{1L}^{-1}, \tag{4}$$

where $B_{1L}^{-1} = \sum_{i=1}^N R_i^T A_i^{-1} R_i$ and the local matrices are defined by $vA_i w = a(v,w)$ for all $v, w \in V_0^h(D_i')$, $i = 1, \ldots, N$ (see [5]).

We denote by $\{y_i\}_{i=1}^{N_v}$ the vertices of the coarse mesh $\mathscr{T}^H$ and define

$$\omega_i = \bigcup \{K \in \mathscr{T}^H; \ \ y_i \in \overline{K}\}, \ \ \omega_K = \bigcup \{\omega_j; \ \ y_j \in \overline{K}\}. \tag{5}$$

Additionally, we use a partition of unity $\{\chi_i\}_{i=1}^{N_v}$ subordinated to the covering $\{\omega_i\}_i^{N_v}$ such that

$$\sum_{i=1}^{N_v} \chi_i = 1, \quad \chi_i \in V^h(D), \quad 0 \le \chi_i \le 1 \quad \text{and} \quad \mathrm{Supp}(\chi_i) \subset \omega_i, i = 1, \ldots, N_v. \tag{6}$$

4 Coarse Space Construction and Dimension Reduction

In this section we define a local spectral multiscale coarse space using eigenvectors of high-anisotropy eigenvalue problems. First we introduce the notation for eigenvalue

problems following [1]. For $i = 1,\dots,N_v$, define the matrix A^{ω_i} and the *modified mass matrix* of same dimension M^{ω_i} by

$$v^T A^{\omega_i} w = \int_{\omega_i} \kappa \nabla v \cdot \nabla w dx \ \text{ and } \ v^T M^{\omega_i} w = \int_{\omega_i} \widetilde{\kappa} v w dx \ \ \forall v, w \in \widetilde{V}^h(\omega_i), \qquad (7)$$

where $\widetilde{V}^h(\omega_i) = \{v \in V^h(\omega_i) : v = 0 \text{ on } \partial\omega_i \cap \partial D\}$. Here $\widetilde{\kappa}$ is an scalar weight derived from the high-anisotropy coefficient matrix $\kappa = [\kappa_{ij}]$ and contains the relevant information we need for the construction of the coarse basis functions. Several possible choices for $\widetilde{\kappa}$ can be considered. Here $\widetilde{\kappa}$ is defined by

$$\widetilde{\kappa} = \max\left\{ \sum_{i=1}^{N} \kappa \nabla \xi_i \cdot \nabla \xi_i, \sum_{j=1}^{N_v} \kappa \nabla \chi_j \cdot \nabla \chi_j \right\}, \qquad (8)$$

where $\{\xi_j\}_{j=1}^N$ and $\{\chi_i\}_{i=1}^{N_v}$ are the partition of unity introduced in (3) and (6), respectively. From now on, we assume that the overlapping decomposition is constructed from the coarse mesh and then $\xi_i = \chi_i$ and $D_i' = \omega_i$ for all $i = 1,\dots,N = N_v$, and $\delta \asymp H$. We consider the finite dimensional symmetric eigenvalue problems $A^{\omega_i}\psi = \widetilde{\lambda} M^{\omega_i}\psi$, with A^{ω_i} and M^{ω_i} defined by (7) and (8), $i = 1,\dots,N$. Denote its eigenvalues and eigenvectors by $\{\widetilde{\lambda}_\ell^{\omega_i}\}$ and $\{\psi_\ell^{\omega_i}\}$, respectively. Note that the eigenvectors $\{\psi_\ell^{\omega_i}\}$ form an orthonormal basis of $\widetilde{V}^h(\omega_i)$ with respect to the M^{ω_i} inner product. Assume that $\widetilde{\lambda}_1^{\omega_i} \le \widetilde{\lambda}_2^{\omega_i} \le \cdots \le \widetilde{\lambda}_\ell^{\omega_i} \le \dots$, and note that $\widetilde{\lambda}_1^{\omega_i} = 0$ for all interior subdomains. In particular, $\psi_\ell^{\omega_i}$ denotes the ℓ-th eigenvector of the matrix associated to the neighborhood of y_i, $i = 1,\dots,N_v$.

Let $\{\chi_i\}_{i=1}^{N_v}$ be a partition of unity (3). Define the coarse basis functions

$$\Phi_{i,\ell} = I^h(\chi_i \psi_\ell^{\omega_i}) \quad \text{for } 1 \le \ell \le L_i \text{ and } 1 \le i \le N_v, \qquad (9)$$

where I^h is the fine-scale nodal value interpolation and L_i is an integer number for each $i = 1,\dots,N_v$. Denote by V_0^H the *spectral multiscale* space

$$V_0^H = \text{span}\{\Phi_{i,\ell} : 1 \le \ell \le L_i \text{ and } 1 \le i \le N_v\}. \qquad (10)$$

The idea is to use only eigenvectors of contrast dependent eigenvalues. Next, we discuss how the choice of $\widetilde{\kappa}$ affects the eigenvalues. If we choose χ_i to be piecewise linear functions on the coarse grid, then, it is easy to see that we have $\widetilde{\kappa}(x_1,x_2) = \sum_i \eta(x_1,x_2)|\partial_{x_1}\chi_i(x_1,x_2)|^2 + |\partial_{x_2}\chi_i(x_1,x_2)|^2$ and $\widetilde{\kappa}$ will have similar behavior as $\eta(x)$. In this case, one can show that the number of small eigenvalues is the same as the fine degrees of freedom in the form of discrete functions that depend on x_2 within high-anisotropy inclusions and channels. Indeed, if we consider the associated Rayleigh quotient, $R(v) = \frac{v^T A^{\omega_i} w}{v^T M^{\omega_i} w}$, we have

$$R(v) = \frac{\int_{\omega_i} \kappa \nabla v \cdot \nabla v}{\int_{\omega_i} \widetilde{\kappa} v^2} = \frac{\int_{\omega_i} \eta(x_1,x_2)|\partial_{x_1} v(x_1,x_2)|^2 + |\partial_{x_2} v(x_1,x_2)|^2}{\int_{\omega_i} \left(\sum_i \eta(x_1,x_2)|\partial_{x_1}\chi_i(x_1,x_2)|^2 + |\partial_{x_2}\chi_i(x_1,x_2)|^2\right) v(x_1,x_2)^2}.$$

Then, for functions that depends only on x_2 inside the region R where η is high, the numerator reduces to $\int_{\omega_i \setminus R} \left(|\partial_{x_1} v(x_1,x_2)|^2 + |\partial_{x_2} v(x_1,x_2)|^2\right) + \int_R |\partial_{x_2} v(x_1,x_2)|^2$

(which is independent of the high value of $\eta(x)$ in R) and the quotient will go to zero as the value of η in R goes to infinity. Including all fine grid functions of x_2 into the coarse space can lead to a high dimensional coarse spaces. Note that the dimension of the coarse space will be much higher than the case with scalar coefficient κ where the number of small eigenvalues is equal to the number of isolated inclusions and channels within a coarse block; see [1, 3]. To reduce the dimension of the coarse space, we propose the use of multiscale basis functions.

We are interested in partition of unity functions that can reduce the number of degrees of freedom associated with isolated high-anisotropy inclusions. This can be achieved by minimizing high-conductivity components for the scalar function $\widetilde{\kappa}$. In particular, by choosing multiscale finite element basis functions or energy minimizing basis functions (e.g., [6]), we can eliminate all isolated high-conductivity inclusions. This can be observed in our numerical experiments. We recall the definition of the "standard" multiscale finite element basis functions that coincide with (the piecewise linear functions on the coarse grid) χ_i^0 on the boundaries of the coarse partition. They are denoted by χ_i^{ms} and satisfy:

$$-\operatorname{div}(\kappa\nabla\chi_i^{ms}) = 0 \ \text{ in } K \in \omega_i, \quad \chi_i^{ms} = \chi_i^0 \ \text{ in } \partial K, \ \forall\, K \in \omega_i, \tag{11}$$

where K is a coarse grid block within ω_i, see [2] for more details and more general multiscale basis functions constructions. In Fig. 1, we depict $\eta(x)$ (left picture) and $\widetilde{\kappa}$ (right picture) using multiscale basis functions on the coarse grid. One can observe that isolated inclusions are removed in $\widetilde{\kappa}$. The coarse space contains functions depending only on x_2 within long channels. The situation is more complicated if high-anisotropy regions form complex channel patterns. For example, if high-anisotropy region is vertical for the coefficients considered in our numerical example, then initial multiscale spaces can represent them and no additional degrees are needed. More complex channel shapes will be studied elsewhere.

We note that for the proposed methods, in each ω_i, $i = 1,\ldots,N_v$, we only need to specify the number of eigenvectors L_i based on the quantities $\{1/\widetilde{\lambda}_l^{\omega_i}\}$. These eigenvectors are used to construct the coarse space. In practice, one only needs to compute the first L_i eigenvalues. Hierarchical approximation with several triangulations can also be considered for the eigenvalues and eigenvectors.

Weighted L^2 approximation and weighted H^1 stability properties of the coarse space V_0^H in (10) hold (as in [1, 3]). In order to describe better these properties of V_0^H, we need to introduce a relevant interpolation operator. Given $v \in V^h(\omega_i)$, set

$$I_{L_i}^{\omega_i} v = \sum_{\ell=1}^{L_i} \left(\int_{\omega_i} \widetilde{\kappa} v \psi_\ell^{\omega_i} dx \right) \psi_\ell^{\omega_i}, \quad i = 1,\ldots,N_v, \tag{12}$$

and define the coarse interpolation $I_0 : V^h(D) \to V_0^H$ by

$$I_0 v = \sum_{i=1}^{N_v} \sum_{\ell=1}^{L_i} \left(\int_{\omega_i} \widetilde{\kappa} v \psi_\ell^{\omega_i} dx \right) I^h(\chi_i \psi_\ell^{\omega_i}) = \sum_{i=1}^{N_v} I^h \left(\chi_i (I_{L_i}^{\omega_i} v) \right), \tag{13}$$

where I^h is the fine-scale nodal value interpolation.

Lemma 1. *For each coarse element K we have*

- $\int_K \widetilde{\kappa}(v - I_0 v)^2 \preceq \widetilde{\lambda}_{K,L+1}^{-1} \int_{\omega_K} \kappa \nabla v \cdot \nabla v dx$
- $\int_K \kappa \nabla I_0 v \cdot \nabla I_0 v dx \preceq \max\{1, \widetilde{\lambda}_{K,L+1}^{-1}\} \int_{\omega_K} \kappa \nabla v \cdot \nabla v dx,$

where $\widetilde{\lambda}_{K,L+1} = \min_{y_i \in K} \widetilde{\lambda}_{L_i+1}^{\omega_i}$ *and* ω_K *is defined in (5).*

Using Lemma 1, we can estimate the condition number of the preconditioned operator $B^{-1}A$ with B^{-1} defined in (4) using the coarse space V_0^H in (10). Following [1, 3], one has the following result.

Theorem 1. *The condition number, cond$(B^{-1}A)$, of the preconditioned operator $B^{-1}A$ with B^{-1} defined in (4) satisfies*

$$cond(B^{-1}A) \preceq 1 + \widetilde{\lambda}_{L+1}^{-1}, \quad \text{where} \quad \widetilde{\lambda}_{L+1} = \min_{1 \le i \le N_v} \widetilde{\lambda}_{L_i+1}^{\omega_i}.$$

Recall that we assumed $\xi_i = \chi_i$, $i = 1, \ldots, N = N_v$. It can be easily shown that if we choose L_i as the number of contrast dependent eigenvalues, then $\widetilde{\lambda}_{L+1}$ scales as $O(1)$, i.e., independent of the contrast. The dependency of the condition number on δ and H is controlled by the partition of unity $\{\chi_i\}$. The condition number is independent of h and it is, in the general case of different partitions of unity, $\{\chi_i\}$ and $\{\xi_i\}$, of order $O(H^2/\delta^2)$, see [3].

5 Numerical Results

In this section, we show representative $2D$ numerical results for the additive preconditioner (4) with the local spectral multiscale coarse space defined in (10). We take $D = [0,1] \times [0,1]$ that is divided into 10×10 equal square coarse blocks to construct the coarse mesh. Inside each coarse block we use a fine-scale triangulation where triangular elements constructed from 10×10 squares are used.

We test our approach on a permeability field that contains inclusions and channels on a background of conductivity one (see the left picture of Fig. 1 for $\eta(x)$ in (1)). We use multiscale finite element basis functions as the initial partition of unity. From the right picture of Fig. 1 we see that the modified weight $\widetilde{\kappa}$ does not contain any isolated inclusions and only contains long high-anisotropy channels connecting boundaries of coarse-grid blocks. This is automatically achieved from the choice of the partition of unity functions. There are fewer small (asymptotically vanishing) eigenvalues when local eigenvalue problem is solved with the modified weight $\widetilde{\kappa}$. Thus, a good choice of partition of unity functions χ_i in (8) will ensure fewer new multiscale basis functions needed to achieve an optimal convergence with respect to the contrast. Numerical results are presented in Table 1. We observe that using the proposed coarse spaces, the number of iterations is independent of contrast. In Table 1 we also show the dimension of the coarse spaces. The dimension of the local spectral coarse space is smaller if we use $\widetilde{\kappa}$ in (10) with multiscale basis functions instead of piecewise linear basis functions.

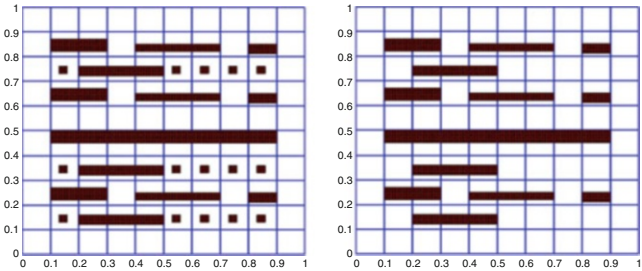


Fig. 1. *Left*: Coarse mesh and coefficient (we plot $\eta(x) = 10^6$ and recall that $\eta(x) = 1$ elsewhere). *Right*: Coefficient $\widetilde{\kappa}$ in (8) using multiscale basis functions (we plot $\widetilde{\kappa}(x) \geq 10^6$). See Table 1

η	LIN	MS	EMF	LSM (bilin. χ_i)	LSM (MS χ_i)
10^3	113(1.48e+2)	122(1.51e+2)	115(1.81e+2)	53(23.21)	55(26.9)
10^4	257(1.35e+3)	258(1.28e+3)	231(9.70e+2)	41(53.63)	28(5.82)
10^5	435(1.34e+4)	483(1.26e+4)	416(9.64e+3)	28(5.642)	29(6.02)
10^6	627(1.34e+5)	709(1.27e+5)	599(9.63e+4)	30(5.753)	29(6.04)
Dim	81=0.79%	81=0.79%	81=0.79%	732=7.19%	497=4.87%

Table 1. Number of iterations and estimated condition number for the PCG and various values of η with the coefficient depicted in Figure 1. We set the tolerance to $1e-10$, $H = 1/10$, $h = 1/100$, and $\dim(V_h) = 10201$. The notation MS stands for the (linear boundary condition) multiscale (MS) coarse space, EMF is the energy minimizing coarse space, see e.g., [6], and LSM is the local spectral multiscale coarse space defined in (10). We select the first L eigenvalues such that $\tilde{\lambda}_L - \tilde{\lambda}_{L-1} > 0.05$ (which is and easy way to select the small eigenvalues- in this example, the value 0.05 was chose by trial-and-error).

6 Discussion on Coarse Space Dimension Reduction

Now we discuss approaches to avoid the use of high-dimensional coarse spaces without sacrificing the efficiency of the preconditioner at the expense of solving problems in high-anisotropy channels. As was observed in the presented numerical tests, the strongly anisotropic channels cause a substantial increase of the size of the coarse space and the complexity of the method. To avoid this, we can replace the coarse solve $R_0^T A_0^{-1} R_0$ in (4) by $R_0^T \widetilde{A}_0^{-1} R_0 + R_{an}^T A_{an}^{-1} R_{an}$. Here the matrix $\widetilde{A}_0$ is a small dimensional coarse matrix. The matrix A_{an} is acting on the fine-mesh degrees restricted to subdomain of high-anisotropy channels Ω_{an}. It is based on the original matrix A and is constructed locally (element-by-element) by preserving the strongest links (off-diagonal entries) of the element stiffness matrices in the channels. To illustrate this idea, which was developed in [4] for Crouzeix-Raviart elements, we write an element stiffness matrix A_e for $e \subset \Omega_{an}$: $A_e = [b_e + c_e,\ -c_e,\ -b_e; -c_e,\ a_e + c_e,\ -a_e; -b_e,\ -a_e,\ a_e + b_e]$, where $|a_e| \leq b_e \leq c_e$. Then the matrix A_{an} is defined as assembly of the matrices $B_e = [c_e,\ -c_e,\ 0; -c_e,\ c_e,\ 0; 0,\ 0,\ 0]$, $e \subset \Omega_{an}$. It is easy

to see that A_{an} is a stiffness matrix corresponding to a diffusion problem defined on a carcass of piecewise linear lines in Ω_{an} following the directions of dominating anisotropy.

In the case of apparent dominant anisotropy direction (i.e., when A_{an} is block diagonal with tridiagonal blocks), inverting A_{an} will involve solving block-diagonal problems with tridiagonal blocks (in 2-D only). In this case optimal complexity is achieved by using a sparse direct solver. In general, one may consider including some of the degrees of freedom associated with high-anisotropy regions into the coarse space while using A_{an}^{-1} to handle the others. Another possibility is to use an auxiliary space of Crouzeix-Raviart elements combined with the technique from [4]. These issues will be studied in our subsequent work.

Acknowledgments The work of all authors has been partially supported by award KUS-C1-016-04, made by King Abdullah University of Science and Technology (KAUST). The work of RL has been supported in part by the US NSF Grant DMS-1016525. The work of SM was partially supported by the Bulgarian NSF Grant DO 02-147/08. The work of YE has been supported by the DOE and NSF (DMS 0934837, DMS 0902552, and DMS 0811180).

Bibliography

[1] Y. Efendiev and J. Galvis. A domain decomposition preconditioner for multiscale high-contrast problems. In Y. Huang, R. Kornhuber, O. Widlund, and J. Xu, editors, *Domain Decomposition Methods in Science and Engineering XIX*, volume 78 of *Lecture Notes in Computational Science and Engineering*, pages 189–196. Springer, 2011.

[2] Yalchin Efendiev and Thomas Y. Hou. *Multiscale finite element methods. Theory and applications*, volume 4 of *Surveys and Tutorials in the Applied Mathematical Sciences*. Springer, New York, 2009.

[3] Juan Galvis and Yalchin Efendiev. Domain decomposition preconditioners for multiscale flows in high contrast media: reduced dimension coarse spaces. *Multiscale Model. Simul.*, 8(5):1621–1644, 2010.

[4] J. Kraus and S. Margenov. *Robust Algebraic Multilevel Methods and Algorithms*, volume 5 of *Radon Series on Comput. Appl. Math.* de Gruyter, 2009.

[5] Andrea Toselli and Olof Widlund. *Domain decomposition methods—algorithms and theory*, volume 34 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin, 2005.

[6] Jinchao Xu and Ludmil Zikatanov. On an energy minimizing basis for algebraic multigrid methods. *Comput. Vis. Sci.*, 7(3-4):121–127, 2004.

A Robust FEM-BEM Solver for Time-Harmonic Eddy Current Problems

Michael Kolmbauer[1] and Ulrich Langer[2]

[1] Institute of Computational Mathematics, Johannes Kepler University, Linz, Austria, kolmbauer@numa.uni-linz.ac.at
[2] RICAM, Austrian Academy of Sciences, Linz, Austria, ulanger@numa.uni-linz.ac.at

Summary. This paper is devoted to the construction and analysis of robust solution techniques for time-harmonic eddy current problems in unbounded domains. We discretize the time-harmonic eddy current equation by means of a symmetrically coupled finite and boundary element method, taking care of the different physical behavior in conducting and non-conducting subdomains, respectively. We construct and analyse a block-diagonal preconditioner for the system of coupled finite and boundary element equations that is robust with respect to the space discretization parameter as well as all involved "bad" parameters like the frequency, the conductivity and the reluctivity. Block-diagonal preconditioners can be used for accelerating iterative solution methods such like the Minimal Residual Method.

1 Introduction

In many practical applications, the excitation is time-harmonic. Switching from the time domain to the frequency domain allows us to replace expensive time-integration procedures by the solution of a system of partial differential equations for the amplitudes belonging to the sine- and to the cosine-excitation. Following this strategy, [7, 13] and [4, 5] applied harmonic and multiharmonic approaches to parabolic initial-boundary value problems and the eddy current problem, respectively. Indeed, in [13], a preconditioned MinRes solver for the solution of the eddy current problem in bounded domains was constructed that is robust with respect to both the discretization parameter h and the frequency ω. The key point of this parameter-robust solver is the construction of a block-diagonal preconditioner, where standard $\mathbf{H}(\mathbf{curl})$ FEM magneto-static problems have to be solved or preconditioned. The aim of this contribution is to generalize these ideas to the case of unbounded domains in terms of a coupled Finite Element (FEM) – Boundary Element (BEM) Method. In this case we are also able to construct a block-diagonal preconditioner, where now standard coupled FEM-BEM $\mathbf{H}(\mathbf{curl})$ problems, as arising in the magneto-static case, have to be solved or preconditioned. We mention, that this preconditioning technique fits into the framework of operator preconditioning, see, e.g. [1, 11, 16, 19].

R. Bank et al. (eds.), *Domain Decomposition Methods in Science and Engineering XX*, Lecture Notes in Computational Science and Engineering 91,
DOI 10.1007/978-3-642-35275-1_34,

The paper is now organized as follows. We introduce the frequency domain equations in Sect. 2. In the same section, we provide the symmetrically coupled FEM-BEM discretization of these equations. In Sect. 3, we construct and analyse our parameter-robust block-diagonal preconditioner used in a MinRes setting for solving the resulting system of linear algebraic equations. Finally, we discuss the practical realization of our preconditioner.

2 Frequency Domain FEM-BEM

As a model problem, we consider the following eddy current problem:

$$\begin{cases} \sigma \frac{\partial \mathbf{u}}{\partial t} + \mathbf{curl}\,(\nu_1\,\mathbf{curl}\,\mathbf{u}) = \mathbf{f} & \text{in } \Omega_1 \times (0,T), \\ \mathbf{curl}\,(\mathbf{curl}\,\mathbf{u}) = \mathbf{0} & \text{in } \Omega_2 \times (0,T), \\ \operatorname{div}\mathbf{u} = 0 & \text{in } \Omega_2 \times (0,T), \\ \mathbf{u} = \mathscr{O}(|\mathbf{x}|^{-1}) & \text{for } |\mathbf{x}| \to \infty, \\ \mathbf{curl}\,\mathbf{u} = \mathscr{O}(|\mathbf{x}|^{-1}) & \text{for } |\mathbf{x}| \to \infty, \\ \mathbf{u} = \mathbf{u_0} & \text{on } \Omega_1 \times \{0\}, \\ \mathbf{u_1} \times \mathbf{n} = \mathbf{u_2} \times \mathbf{n} & \text{on } \Gamma \times (0,T), \\ \nu_1 \mathbf{curl}\,\mathbf{u_1} \times \mathbf{n} = \mathbf{curl}\,\mathbf{u_2} \times \mathbf{n} & \text{on } \Gamma \times (0,T), \end{cases} \tag{1}$$

where the computational domain $\Omega = \mathbb{R}^3$ is split into the two non-overlapping subdomains Ω_1 and Ω_2. The conducting subdomain Ω_1 is assumed to be a simply connected Lipschitz polyhedron, whereas the non-conducting subdomain Ω_2 is the complement of Ω_1 in $\mathbb{R}^3$, i.e $\mathbb{R}^3 \setminus \overline{\Omega}_1$. Furthermore, we denote by Γ the interface between the two subdomains, i.e. $\Gamma = \overline{\Omega}_1 \cap \overline{\Omega}_2$. The exterior unit normal vector of Ω_1 on Γ is denoted by $\mathbf{n}$, i.e. $\mathbf{n}$ points from Ω_1 to Ω_2. The reluctivity ν_1 is supposed to be independent of $|\mathbf{curl}\,\mathbf{u}|$, i.e. we assume the eddy current problem (1) to be linear. The conductivity σ is zero in Ω_2, and piecewise constant and uniformly positive in Ω_1.

We assume, that the source $\mathbf{f}$ is given by a time-harmonic excitation with the frequency $\omega > 0$ and amplitudes $\mathbf{f^c}$ and $\mathbf{f^s}$ in the conducting domain Ω_1. Therefore, the solution $\mathbf{u}$ is time-harmonic as well, with the same base frequency ω, i.e.

$$\mathbf{u}(\mathbf{x},t) = \mathbf{u^c}(\mathbf{x})\cos(\omega t) + \mathbf{u^s}(\mathbf{x})\sin(\omega t). \tag{2}$$

In fact, (2) is the real reformulation of a complex time-harmonic approach $\mathbf{u}(\mathbf{x},t) = \hat{\mathbf{u}}(\mathbf{x})e^{i\omega t}$ with the complex-valued amplitude $\hat{\mathbf{u}} = \mathbf{u^c} - i\mathbf{u^s}$. Using the time-harmonic representation (2) of the solution, we can state the eddy current problem (1) in the frequency domain as follows:

$$\text{Find } \mathbf{u} = (\mathbf{u^c}, \mathbf{u^s}): \quad \begin{cases} \omega\sigma\,\mathbf{u^s} + \mathbf{curl}\,(\nu_1\,\mathbf{curl}\,\mathbf{u^c}) = \mathbf{f^c} & \text{in } \Omega_1, \\ \mathbf{curl}\,\mathbf{curl}\,\mathbf{u^c} = \mathbf{0} & \text{in } \Omega_2, \\ -\omega\sigma\,\mathbf{u^c} + \mathbf{curl}\,(\nu_1\,\mathbf{curl}\,\mathbf{u^s}) = \mathbf{f^s} & \text{in } \Omega_1, \\ \mathbf{curl}\,\mathbf{curl}\,\mathbf{u^s} = \mathbf{0} & \text{in } \Omega_2, \end{cases} \tag{3}$$

with the corresponding decay and interface conditions from (1).

Remark 1. In practice, the reluctivity ν_1 depends on the inductivity $|\mathbf{curl}\,\mathbf{u}|$ in a nonlinear way in ferromagnetic materials. Having in mind applications to problems with nonlinear reluctivity, we prefer to use the real reformulation (3) instead of a complex approach. For overcoming the nonlinearity the preferable way is to apply Newton's method due to its fast convergence. It turns out, that Newton's method cannot be applied to the nonlinear complex-valued system (see [4]), but it can be applied to the reformulated real-valued system. Anyhow, the analysis of the linear problem also helps to construct efficient solvers for the nonlinear problem.

Deriving the variational formulation and integrating by parts once more in the exterior domain yields: Find $(\mathbf{u^c},\mathbf{u^s}) \in \mathbf{H}(\mathbf{curl},\Omega_1)^2$ such that

$$\begin{cases} \omega(\sigma\mathbf{u^s},\mathbf{v^c})_{L_2(\Omega_1)} + (\nu_1\mathbf{curl}\,\mathbf{u^c},\mathbf{curl}\,\mathbf{v^c})_{L_2(\Omega_1)} - \langle\gamma_N\mathbf{u^c},\gamma_D\mathbf{v^c}\rangle_\tau = \langle\mathbf{f^c},\mathbf{v^c}\rangle, \\ -\omega(\sigma\mathbf{u^c},\mathbf{v^s})_{L_2(\Omega_1)} + (\nu_1\mathbf{curl}\,\mathbf{u^s},\mathbf{curl}\,\mathbf{v^s})_{L_2(\Omega_1)} - \langle\gamma_N\mathbf{u^s},\gamma_D\mathbf{v^s}\rangle_\tau = \langle\mathbf{f^s},\mathbf{v^s}\rangle, \end{cases}$$

for all $(\mathbf{v^c},\mathbf{v^s}) \in \mathbf{H}(\mathbf{curl},\Omega_1)^2$. Here γ_D and γ_N denote the Dirichlet trace $\gamma_D := \mathbf{n}\times(\mathbf{u}\times\mathbf{n})$ and the Neumann trace $\gamma_N := \mathbf{curl}\,\mathbf{u}\times\mathbf{n}$ on the interface Γ. $\langle\cdot,\cdot\rangle_\tau$ denotes the $L_2(\Gamma)$-based duality product. In order to deal with the expression on the interface Γ, we use the framework of the symmetric FEM-BEM coupling for eddy current problems (see [10]). So, using the boundary integral operators $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ and $\mathbf{N}$, as defined in [10], we end up with the weak formulation of the time-harmonic eddy current problem: Find $(\mathbf{u^c},\mathbf{u^s}) \in \mathbf{H}(\mathbf{curl},\Omega_1)^2$ and $(\lambda^\mathbf{c},\lambda^\mathbf{s}) \in \mathbf{H}_{\|}^{-\frac{1}{2}}(\mathrm{div}_\Gamma 0,\Gamma)^2$ such that

$$\begin{cases} \omega(\sigma\mathbf{u^s},\mathbf{v^c})_{L_2(\Omega_1)} + (\nu_1\mathbf{curl}\,\mathbf{u^c},\mathbf{curl}\,\mathbf{v^c})_{L_2(\Omega_1)}, \\ \qquad -\langle\mathbf{N}(\gamma_D\mathbf{u^c}),\gamma_D\mathbf{v^c}\rangle_\tau + \langle\mathbf{B}(\lambda^\mathbf{c}),\gamma_D\mathbf{v^c}\rangle_\tau = \langle\mathbf{f^c},\mathbf{v^c}\rangle, \\ \qquad \langle\mu^\mathbf{c},(\mathbf{C}-\mathbf{Id})(\gamma_D\mathbf{u^c})\rangle_\tau - \langle\mu^\mathbf{c},\mathbf{A}(\lambda^\mathbf{c})\rangle_\tau = 0, \\ -\omega(\sigma\mathbf{u^c},\mathbf{v^s})_{L_2(\Omega_1)} + (\nu_1\mathbf{curl}\,\mathbf{u^s},\mathbf{curl}\,\mathbf{v^s})_{L_2(\Omega_1)}, \\ \qquad -\langle\mathbf{N}(\gamma_D\mathbf{u^s}),\gamma_D\mathbf{v^s}\rangle_\tau + \langle\mathbf{B}(\lambda^\mathbf{s}),\gamma_D\mathbf{v^s}\rangle_\tau = \langle\mathbf{f^s},\mathbf{v^s}\rangle, \\ \qquad \langle\mu^\mathbf{s},(\mathbf{C}-\mathbf{Id})(\gamma_D\mathbf{u^s})\rangle_\tau - \langle\mu^\mathbf{s},\mathbf{A}(\lambda^\mathbf{s})\rangle_\tau = 0, \end{cases} \tag{4}$$

for all $(\mathbf{v^c},\mathbf{v^s}) \in \mathbf{H}(\mathbf{curl},\Omega_1)^2$ and $(\mu^\mathbf{c},\mu^\mathbf{s}) \in \mathbf{H}_{\|}^{-\frac{1}{2}}(\mathrm{div}_\Gamma 0,\Gamma)^2$. This variational form is the starting point of the discretization in space. Therefore, we use a regular triangulation $\mathscr{T}_h$, with mesh size $h > 0$, of the domain Ω_1 with tetrahedral elements. $\mathscr{T}_h$ induces a mesh $\mathscr{K}_h$ of triangles on the boundary Γ. On these meshes, we consider Nédélec basis functions of order p yielding the conforming finite element subspace $\mathscr{ND}_p(\mathscr{T}_h)$ of $\mathbf{H}(\mathbf{curl},\Omega_1)$, see [17]. Further, we use the space of divergence free Raviart-Thomas basis functions $\mathscr{RT}_p^0(\mathscr{K}_h) := \{\lambda_h \in \mathscr{RT}_p(\mathscr{K}_h), \mathrm{div}_\Gamma\lambda_h = 0\}$ being a conforming finite element subspace of $\mathbf{H}_{\|}^{-\frac{1}{2}}(\mathrm{div}_\Gamma 0,\Gamma)$. Let $\{\varphi_i\}$ denote the basis of $\mathscr{ND}_p(\mathscr{T}_h)$, and let $\{\psi_i\}$ denote the basis of $\mathscr{RT}_p^0(\mathscr{K}_h)$. Then the matrix entries corresponding to the operators in (4) are given by the formulas

$$
\begin{aligned}
(\mathbf{K})_{ij} &:= (\nu\ \mathbf{curl}\,\varphi_{\mathbf{i}}, \mathbf{curl}\,\varphi_{\mathbf{j}})_{\mathbf{L}_2(\Omega_1)} - \langle \mathbf{N}(\gamma_D\varphi_{\mathbf{i}}), \gamma_D\varphi_{\mathbf{j}}\rangle_\tau,\\
(\mathbf{M})_{ij} &:= \omega(\sigma\varphi_{\mathbf{i}}, \varphi_{\mathbf{j}})_{\mathbf{L}_2(\Omega_1)},\\
(\mathbf{A})_{ij} &:= \langle \psi_{\mathbf{i}}, \mathbf{A}(\psi_{\mathbf{j}})\rangle_\tau,\\
(\mathbf{B})_{ij} &:= \langle \psi_{\mathbf{i}}, (\mathbf{C}-\mathbf{Id})(\gamma_D\varphi_{\mathbf{j}})\rangle_\tau.
\end{aligned}
$$

The entries of the right-hand side vector are given by the formulas $(\mathbf{f}^{\mathbf{c}})_i := (\mathbf{f}^{\mathbf{c}}, \varphi_{\mathbf{i}})_{\mathbf{L}_2(\Omega_1)}$ and $(\mathbf{f}^{\mathbf{s}})_i := (\mathbf{f}^{\mathbf{s}}, \varphi_{\mathbf{i}})_{\mathbf{L}_2(\Omega_1)}$. The resulting system $\mathscr{A}\,\mathbf{x} = \mathbf{f}$ of the coupled finite and boundary element equations has now the following structure:

$$
\begin{pmatrix}
\mathbf{M} & 0 & \mathbf{K} & \mathbf{B}^T\\
0 & 0 & \mathbf{B} & -\mathbf{A}\\
\mathbf{K} & \mathbf{B}^T & -\mathbf{M} & 0\\
\mathbf{B} & -\mathbf{A} & 0 & 0
\end{pmatrix}
\begin{pmatrix}
\mathbf{u}^{\mathbf{s}}\\ \lambda^{\mathbf{s}}\\ \mathbf{u}^{\mathbf{c}}\\ \lambda^{\mathbf{c}}
\end{pmatrix}
=
\begin{pmatrix}
\mathbf{f}^{\mathbf{c}}\\ 0\\ \mathbf{f}^{\mathbf{s}}\\ 0
\end{pmatrix}. \tag{5}
$$

In fact, the system matrix $\mathscr{A}$ is symmetric and indefinite and obtains a double saddle-point structure. Since $\mathscr{A}$ is symmetric, the system can be solved by a Min-Res method, see, e.g., [18]. Anyhow, the convergence rate of any iterative method deteriorates with respect to the meshsize h and the "bad" parameters ω, ν and σ, if applied to the unpreconditioned system (5). Therefore, preconditioning is a challenging topic.

3 A Parameter-Robust Preconditioning Technique

In this section, we investigate a preconditioning technique for double saddle-point equations with the block-structure (5). Due to the symmetry and coercivity properties of the underlying operators, the blocks fulfill the following properties: $\mathbf{K} = \mathbf{K}^T \geq 0$, $\mathbf{M} = \mathbf{M}^T > 0$ and $\mathbf{A} = \mathbf{A}^T > 0$.

In [19] a parameter-robust block-diagonal preconditioner for the distributed optimal control of the Stokes equations is constructed. The structural similarities to that preconditioner gives us a hint how to choose the block-diagonal preconditioner in our case. Therefore, we propose the following preconditioner

$$
\mathscr{C} = \mathrm{diag}\ (\mathscr{I}_{FEM}, \mathscr{I}_{BEM}, \mathscr{I}_{FEM}, \mathscr{I}_{BEM}),
$$

where the diagonal blocks are given by $\mathscr{I}_{FEM} = \mathbf{M} + \mathbf{K}$ and $\mathscr{I}_{BEM} = \mathbf{A} + \mathbf{B}\mathscr{I}_{FEM}^{-1}\mathbf{B}^T$. Being aware that $\mathscr{I}_{FEM}$ and $\mathscr{I}_{BEM}$ are symmetric and positive definite, we conclude that $\mathscr{C}$ is also symmetric and positive definite. Therefore, $\mathscr{C}$ induces the energy norm $\|\mathbf{u}\|_{\mathscr{C}} = \sqrt{\mathbf{u}^T\mathscr{C}\mathbf{u}}$. Using this special norm, we can apply the Theorem of Babuška-Aziz [3] to the variational problem:

$$
\text{Find } \mathbf{x} \in \mathbb{R}^N: \quad \mathbf{w}^T\mathscr{A}\mathbf{x} = \mathbf{w}^T\mathbf{f}, \quad \forall \mathbf{w} \in \mathbb{R}^N.
$$

The main result is now summarized in the following lemma.

Lemma 1. *The matrix $\mathscr{A}$ satisfies the following norm equivalence inequalities:*

$$\frac{1}{\sqrt{7}}\|\mathbf{x}\|_{\mathscr{C}} \leq \sup_{\mathbf{w}\neq 0}\frac{\mathbf{w}^T\mathscr{A}\mathbf{x}}{\|\mathbf{w}\|_{\mathscr{C}}} \leq 2\|\mathbf{x}\|_{\mathscr{C}} \quad \forall \mathbf{x}\in\mathbb{R}^N.$$

Proof. Throughout the proof, we use the following notation: $\mathbf{x} = (\mathbf{x_1},\mathbf{x_2},\mathbf{x_3},\mathbf{x_4})^T$ and $\mathbf{y} = (\mathbf{y_1},\mathbf{y_2},\mathbf{y_3},\mathbf{y_4})^T$. The upper bound follows by reapplication of Cauchy's inequality several time. The expressions corresponding to the Schur complement can be derived in the following way:

$$\mathbf{y_1}^T\mathbf{B}^T\mathbf{x_4} = \mathbf{y_1}\mathscr{I}_{FEM}^{1/2}\mathscr{I}_{FEM}^{-1/2}\mathbf{B}^T\mathbf{x_4} \leq \|\mathscr{I}_{FEM}^{1/2}\mathbf{y_1}\|_{l_2}\|\mathscr{I}_{FEM}^{-1/2}\mathbf{B}^T\mathbf{x_4}\|_{l_2}.$$

Therefore, we end up with an upper bound with constant 2.

In order to compute the lower bound, we use a linear combination of special test vectors. For the choice $\mathbf{w_1} = (\mathbf{x_1},\mathbf{x_2},-\mathbf{x_3},-\mathbf{x_4})^T$, we obtain

$$\mathbf{w_1}^T\mathscr{A}\mathbf{x} = \mathbf{x_1}^T\mathbf{M}\mathbf{x_1} + \mathbf{x_3}^T\mathbf{M}\mathbf{x_3};$$

for $\mathbf{w_2} = (\mathbf{x_3},-\mathbf{x_4},\mathbf{x_1},-\mathbf{x_2})^T$, we get

$$\mathbf{w_2}^T\mathscr{A}\mathbf{x} = \mathbf{x_1}^T\mathbf{K}\mathbf{x_1} + \mathbf{x_3}^T\mathbf{K}\mathbf{x_3} + \mathbf{x_2}^T\mathbf{A}\mathbf{x_2} + \mathbf{x_4}^T\mathbf{A}\mathbf{x_4};$$

for $\mathbf{w_3} = ((\mathbf{x_4}^T\mathbf{B}(\mathbf{K}+\mathbf{M})^{-1})^T,\mathbf{0},(\mathbf{x_2}^T\mathbf{B}(\mathbf{K}+\mathbf{M})^{-1})^T,\mathbf{0})^T$, we have

$$\begin{aligned}\mathbf{w_3}^T\mathscr{A}\mathbf{x} = {}& \mathbf{x_4}^T\mathbf{B}(\mathbf{K}+\mathbf{M})^{-1}\mathbf{B}^T\mathbf{x_4} + \mathbf{x_2}^T\mathbf{B}(\mathbf{K}+\mathbf{M})^{-1}\mathbf{B}^T\mathbf{x_2}\\ &+\mathbf{x_4}^T\mathbf{B}(\mathbf{K}+\mathbf{M})^{-1}\mathbf{M}\mathbf{x_1} + \mathbf{x_4}^T\mathbf{B}(\mathbf{K}+\mathbf{M})^{-1}\mathbf{K}\mathbf{x_3}\\ &+\mathbf{x_2}^T\mathbf{B}(\mathbf{K}+\mathbf{M})^{-1}\mathbf{K}\mathbf{x_1} - \mathbf{x_2}^T\mathbf{B}(\mathbf{K}+\mathbf{M})^{-1}\mathbf{M}\mathbf{x_3};\end{aligned}$$

for $\mathbf{w_4} = (-(\mathbf{x_3}^T\mathbf{K}(\mathbf{K}+\mathbf{M})^{-1})^T,\mathbf{0},-(\mathbf{x_1}^T\mathbf{K}(\mathbf{K}+\mathbf{M})^{-1})^T,\mathbf{0})^T$, we get

$$\begin{aligned}\mathbf{w_4}^T\mathscr{A}\mathbf{x} = {}& -\mathbf{x_3}^T\mathbf{K}(\mathbf{K}+\mathbf{M})^{-1}\mathbf{M}\mathbf{x_1} - \mathbf{x_3}^T\mathbf{K}(\mathbf{K}+\mathbf{M})^{-1}\mathbf{K}\mathbf{x_3}\\ &-\mathbf{x_3}^T\mathbf{K}(\mathbf{K}+\mathbf{M})^{-1}\mathbf{B}^T\mathbf{x_4} - \mathbf{x_1}^T\mathbf{K}(\mathbf{K}+\mathbf{M})^{-1}\mathbf{K}\mathbf{x_1}\\ &-\mathbf{x_1}^T\mathbf{K}(\mathbf{K}+\mathbf{M})^{-1}\mathbf{B}^T\mathbf{x_2} + \mathbf{x_1}^T\mathbf{K}(\mathbf{K}+\mathbf{M})^{-1}\mathbf{M}\mathbf{x_3};\end{aligned}$$

and, finally, for the choice $\mathbf{w_5} = (-(\mathbf{x_1}^T\mathbf{M}(\mathbf{K}+\mathbf{M})^{-1})^T,\mathbf{0},(\mathbf{x_3}^T\mathbf{M}(\mathbf{K}+\mathbf{M})^{-1})^T,\mathbf{0})^T$, we obtain

$$\begin{aligned}\mathbf{w_5}^T\mathscr{A}\mathbf{x} = {}& -\mathbf{x_1}^T\mathbf{M}(\mathbf{K}+\mathbf{M})^{-1}\mathbf{M}\mathbf{x_1} - \mathbf{x_1}^T\mathbf{M}(\mathbf{K}+\mathbf{M})^{-1}\mathbf{K}\mathbf{x_3}\\ &-\mathbf{x_1}^T\mathbf{M}(\mathbf{K}+\mathbf{M})^{-1}\mathbf{B}^T\mathbf{x_4} + \mathbf{x_3}^T\mathbf{M}(\mathbf{K}+\mathbf{M})^{-1}\mathbf{K}\mathbf{x_1}\\ &+\mathbf{x_3}^T\mathbf{M}(\mathbf{K}+\mathbf{M})^{-1}\mathbf{B}^T\mathbf{x_2} - \mathbf{x_3}^T\mathbf{M}(\mathbf{K}+\mathbf{M})^{-1}\mathbf{M}\mathbf{x_3}.\end{aligned}$$

Therefore, we end up with the following expression

$$\begin{aligned}(\mathbf{w_1}+\mathbf{w_2}+\mathbf{w_3}+\mathbf{w_4}+\mathbf{w_5})^T\mathscr{A}\mathbf{x} &= \mathbf{x_1}^T\mathbf{M}\mathbf{x_1}+\mathbf{x_3}^T\mathbf{M}\mathbf{x_3}\\ &+\mathbf{x_1}^T\mathbf{K}\mathbf{x_1}+\mathbf{x_3}^T\mathbf{K}\mathbf{x_3}+\mathbf{x_2}^T\mathbf{A}\mathbf{x_2}+\mathbf{x_4}^T\mathbf{A}\mathbf{x_4}\\ &+\mathbf{x_4}^T\mathbf{B}(\mathbf{K}+\mathbf{M})^{-1}\mathbf{B}^T\mathbf{x_4}+\mathbf{x_2}^T\mathbf{B}(\mathbf{K}+\mathbf{M})^{-1}\mathbf{B}^T\mathbf{x_2}\\ &-\mathbf{x_3}^T\mathbf{K}(\mathbf{K}+\mathbf{M})^{-1}\mathbf{K}\mathbf{x_3}-\mathbf{x_1}^T\mathbf{K}(\mathbf{K}+\mathbf{M})^{-1}\mathbf{K}\mathbf{x_1}\\ &-\mathbf{x_3}^T\mathbf{M}(\mathbf{K}+\mathbf{M})^{-1}\mathbf{M}\mathbf{x_3}-\mathbf{x_1}^T\mathbf{M}(\mathbf{K}+\mathbf{M})^{-1}\mathbf{M}\mathbf{x_1}\\ &-2\mathbf{x_3}^T\mathbf{K}(\mathbf{K}+\mathbf{M})^{-1}\mathbf{M}\mathbf{x_1}+2\mathbf{x_1}^T\mathbf{K}(\mathbf{K}+\mathbf{M})^{-1}\mathbf{M}\mathbf{x_3}.\end{aligned}$$

For estimating the non-symmetric terms, we use the following result:

$$\begin{aligned}-2\mathbf{x_3}^T\mathbf{K}(\mathbf{K}+\mathbf{M})^{-1}\mathbf{M}\mathbf{x_1} &\geq -2\|(\mathbf{K}+\mathbf{M})^{-1/2}\mathbf{K}\mathbf{x_3}\|_{l_2}\|(\mathbf{K}+\mathbf{M})^{-1/2}\mathbf{M}\mathbf{x_1}\|_{l_2}\\ &\geq -\|(\mathbf{K}+\mathbf{M})^{-1/2}\mathbf{K}\mathbf{x_3}\|_{l_2}^2-\|(\mathbf{K}+\mathbf{M})^{-1/2}\mathbf{M}\mathbf{x_1}\|_{l_2}^2\\ &= -\mathbf{x_3}^T\mathbf{K}(\mathbf{K}+\mathbf{M})^{-1}\mathbf{K}\mathbf{x_3}-\mathbf{x_1}^T\mathbf{M}(\mathbf{K}+\mathbf{M})^{-1}\mathbf{M}\mathbf{x_1}.\end{aligned}$$

Analogously, we obtain

$$2\mathbf{x_1}^T\mathbf{K}(\mathbf{K}+\mathbf{M})^{-1}\mathbf{M}\mathbf{x_3}\geq -\mathbf{x_1}^T\mathbf{K}(\mathbf{K}+\mathbf{M})^{-1}\mathbf{K}\mathbf{x_1}-\mathbf{x_3}^T\mathbf{M}(\mathbf{K}+\mathbf{M})^{-1}\mathbf{M}\mathbf{x_3}.$$

Hence, putting all terms together, we have

$$\begin{aligned}(\mathbf{w_1}+\mathbf{w_2}+\mathbf{w_3}+\mathbf{w_4}+\mathbf{w_5})^T\mathscr{A}\mathbf{x} &\geq \mathbf{x}^T\mathscr{C}\mathbf{x}\\ &-2\mathbf{x_3}^T\mathbf{K}(\mathbf{K}+\mathbf{M})^{-1}\mathbf{K}\mathbf{x_3}-2\mathbf{x_1}^T\mathbf{K}(\mathbf{K}+\mathbf{M})^{-1}\mathbf{K}\mathbf{x_1}\\ &-2\mathbf{x_3}^T\mathbf{M}(\mathbf{K}+\mathbf{M})^{-1}\mathbf{M}\mathbf{x_3}-2\mathbf{x_1}^T\mathbf{M}(\mathbf{K}+\mathbf{M})^{-1}\mathbf{M}\mathbf{x_1}.\end{aligned}$$

In order to get rid of the four remaining terms, we use, for $i=1,3$,

$$\mathbf{x_i}^T\mathbf{K}(\mathbf{K}+\mathbf{M})^{-1}\mathbf{K}\mathbf{x_i}\leq\mathbf{x_i}^T\mathbf{K}\mathbf{x_i}\quad\text{and}\quad\mathbf{x_i}^T\mathbf{M}(\mathbf{K}+\mathbf{M})^{-1}\mathbf{M}\mathbf{x_i}\leq\mathbf{x_i}^T\mathbf{M}\mathbf{x_i}.$$

Hence by adding $\mathbf{w_1}$ and $\mathbf{w_2}$ twice more, we end up with the desired result

$$\underbrace{(3\mathbf{w_1}+3\mathbf{w_2}+\mathbf{w_3}+\mathbf{w_4}+\mathbf{w_5})^T}_{:=\mathbf{w}^T}\mathscr{A}\mathbf{x}\geq\mathbf{x}^T\mathscr{C}\mathbf{x}+\mathbf{x_2}^T\mathbf{A}\mathbf{x_2}+\mathbf{x_4}^T\mathbf{A}\mathbf{x_4}\geq\mathbf{x}^T\mathscr{C}\mathbf{x}.$$

The next step is to compute (and estimate) the $\mathscr{C}$ norm of the special test vector. Straightforward estimations yield

$$\|\mathbf{w}\|_{\mathscr{C}}^2=\|3\mathbf{w_1}+3\mathbf{w_2}+\mathbf{w_3}+\mathbf{w_4}+\mathbf{w_5}\|_{\mathscr{C}}^2\leq 7\,\|\mathbf{x}\|_{\mathscr{C}}^2.$$

This completes the proof.

Now, from Lemma 1, we obtain that the condition number of the preconditioned system can be estimated by the constant $c=2\sqrt{7}$ that is obviously independent of the meshsize h and all involved parameters ω, ν and σ, i.e.

$$\kappa_{\mathscr{C}}(\mathscr{C}^{-1}\mathscr{A}):=\|\mathscr{C}^{-1}\mathscr{A}\|_{\mathscr{C}}\|\mathscr{A}^{-1}\mathscr{C}\|_{\mathscr{C}}\leq 2\sqrt{7}. \tag{6}$$

The condition number defines the convergence behaviour of the MinRes method applied to the preconditioned system (see e.g. [9]), as stated in the following theorem:

Theorem 1 (Robust solver). *The MinRes method applied to the preconditioned system* $\mathscr{C}^{-1}\mathscr{A}\mathbf{u} = \mathscr{C}^{-1}\mathbf{f}$ *converges. At the 2m-th iteration, the preconditioned residual* $\mathbf{r^m} = \mathscr{C}^{-1}\mathbf{f} - \mathscr{C}^{-1}\mathscr{A}\mathbf{u^m}$ *is bounded as*

$$\left\|\mathbf{r^{2m}}\right\|_{\mathscr{C}} \leq \frac{2q^m}{1+q^{2m}} \left\|\mathbf{r^0}\right\|_{\mathscr{C}}, \quad \textit{where} \quad q = \frac{2\sqrt{7}-1}{2\sqrt{7}+1}. \tag{7}$$

4 Conclusion, Outlook and Acknowledgments

The method developed in this work shows great potential for solving time-harmonic eddy current problems in an unbounded domain in a robust way. The solution of a fully coupled 4×4 block-system can be reduced to the solution of a block-diagonal matrix, where each block corresponds to standard problems. We mention, that by analogous procedure, we can state another robust block-diagonal preconditioner $\tilde{\mathscr{C}} = \text{diag}\,(\tilde{\mathscr{I}}_{FEM}, \tilde{\mathscr{I}}_{BEM}, \tilde{\mathscr{I}}_{FEM}, \tilde{\mathscr{I}}_{BEM})$, with $\tilde{\mathscr{I}}_{FEM} = \mathbf{M} + \mathbf{K} + \mathbf{B}^T \tilde{\mathscr{I}}_{BEM}^{-1} \mathbf{B}$ and $\tilde{\mathscr{I}}_{BEM} = \mathbf{A}$, leading to a condition number bound of 4, see e.g. [15].

Of course this block-diagonal preconditioner is only a theoretical one, since the exact solution of the diagonal blocks corresponding to a standard FEM discretized stationary problem and the Schur-complement of a standard FEM-BEM discretized stationary problem are still prohibitively expensive. Nevertheless, as for the FEM discretized version in [13], this theoretical preconditioner allows us replace the solution of a time-dependent problem by the solution of a sequence of time-independent problems in a robust way, i.e. independent of the space and time discretization parameters h and ω and all additional "bad" parameters. Therefore, the issue of finding robust solvers for the fully coupled time-harmonic system matrix $\mathscr{A}$ can be reduced to finding robust solvers for the blocks $\mathscr{I}_{FEM}$ and $\mathscr{I}_{BEM}$, or $\tilde{\mathscr{I}}_{FEM}$ and $\tilde{\mathscr{I}}_{BEM}$. By replacing these diagonal blocks by standard preconditioners, it is straight-forward to derive mesh-independent convergence rates, see, e.g., [8]. Unfortunately, the construction of fully robust preconditioners for the diagonal blocks is not straightforward and has to be studied. Candidates are $\mathscr{H}$ matrix, multigrid multigrid and domain decomposition preconditioners, see, e.g. [2, 6] and [12], respectively.

The preconditioned MinRes solver presented in this paper can also be generalized to eddy current optimal control problems studied in [14] for the pure FEM case in bounded domains.

The authors gratefully acknowledge the financial support by the Austrian Science Fund (FWF) under the grants P19255 and W1214-N15, project DK04. Furthermore, the authors also thank the Austria Center of Competence in Mechatronics (ACCM), which is a part of the COMET K2 program of the Austrian Government, for supporting our work on eddy current problems.

Bibliography

[1] D. N. Arnold, R. S. Falk, and R. Winther. Preconditioning in $H(\text{div})$ and applications. *Math. Comp.*, 66(219):957–984, 1997.

[2] D. N. Arnold, R. S. Falk, and R. Winther. Multigrid in $H(\mathrm{div})$ and $H(\mathrm{curl})$. *Numer. Math.*, 85(2):197–217, 2000.
[3] I. Babuška. Error-bounds for finite element method. *Numer. Math.*, 16(4): 322–333, 1971.
[4] F. Bachinger, U. Langer, and J. Schöberl. Numerical analysis of nonlinear multiharmonic eddy current problems. *Numer. Math.*, 100(4):593–616, 2005.
[5] F. Bachinger, U. Langer, and J. Schöberl. Efficient solvers for nonlinear time-periodic eddy current problems. *Comput. Vis. Sci.*, 9(4):197–207, 2006.
[6] M. Bebendorf. *Hierarchical Matrices*. Springer, 2008.
[7] D. Copeland, M. Kolmbauer, and U. Langer. Domain decomposition solvers for frequency-domain finite element equation. In *Domain Decomposition Methods in Science and Engineering XIX*, volume 78 of *Lect. Notes Comput. Sci. Eng.*, pages 301–308, Heidelberg, 2011. Springer.
[8] S. A. Funken and E. P. Stephan. Fast solvers with block-diagonal preconditioners for linear FEM-BEM coupling. *Numer. Linear Algebra Appl.*, 16(5): 365–395, 2009.
[9] A. Greenbaum. *Iterative methods for solving linear systems*, volume 17 of *Frontiers in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1997.
[10] R. Hiptmair. Symmetric coupling for eddy current problems. *SIAM J. Numer. Anal.*, 40(1):41–65, 2002.
[11] R. Hiptmair. Operator preconditioning. *Comput. Math. Appl.*, 52(5):699–706, 2006.
[12] Q. Hu and J. Zou. A nonoverlapping domain decomposition method for Maxwell's equations in three dimensions. *SIAM J. Numer. Anal.*, 41(5):1682–1708, 2003.
[13] M. Kolmbauer and U. Langer. A frequency-robust solver for the time-harmonic eddy current problem. In *Scientific Computing in Electrical Engineering SCEE 2010*, volume 16 of *Mathematics in Industry*, pages 97–107. Springer, 2011.
[14] M. Kolmbauer and U. Langer. A robust preconditioned Minres-solver for distributed time-periodic eddy current optimal control problems. *SIAM J. Sci. Comput.*, 34(6):B785–B809, 2012.
[15] M. Kolmbauer and U. Langer. A robust preconditioned Minres solver for distributed time-periodic eddy current optimal control problems. Number 28/2011 in Oberwolfach Reports, Oberwolfach, 2011. Mathematisches Forschungsinstitut Oberwolfach: *workshop: Schnelle Löser für Partielle Differentialgleichungen* (May 22nd – May 28th, 2011).
[16] K. A. Mardal and R. Winther. Preconditioning discretizations of systems of partial differential equations. *Numer. Linear Algebra Appl.*, 18(1):1–40, 2011.
[17] J.-C. Nédélec. A new family of mixed finite elements in $\mathbf{R}^3$. *Numer. Math.*, 50 (1):57–81, 1986.
[18] C. C. Paige and M. A. Saunders. Solutions of sparse indefinite systems of linear equations. *SIAM J. Numer. Anal.*, 12(4):617–629, 1975.
[19] W. Zulehner. Nonstandard norms and robust estimates for saddle point problems. *SIAM J. Matrix Anal. Appl.*, 32:536–560, 2011.

Domain Decomposition Methods for Auxiliary Linear Problems of an Elliptic Variational Inequality

Jungho Lee

Argonne National Laboratory, Mathematics and Computer Science Division
julee@mcs.anl.gov

Summary. Elliptic variational inequalities with multiple bodies are considered. It is assumed that an active set method is used to handle the nonlinearity of the inequality constraint, which results in auxiliary linear problems. We describe two domain decomposition methods for solving such linear problems, namely, the FETI-FETI (finite element tearing and interconnecting) and hybrid methods, which are combinations of already existing domain decomposition methods.

Estimates of the condition numbers of both methods are provided. The FETI-FETI method has a condition number which depends linearly on the number of subdomains across each body and polylogarithmically on the number of element across each subdomain. The hybrid method is a scalable alternative to the FETI-FETI method, and has a condition number with two polylogarithmic factors depending on the number of elements across each subdomain and across each body. We present numerical results confirming these theoretical findings.

1 Introduction

Consider the following inequality constrained minimization problem,

$$\begin{aligned} \min \quad & \sum_{i=1}^{N} \left(\frac{1}{2} \int_{\Omega_i} \rho(x) |\nabla u^i(x)|^2 dx - \int_{\Omega_i} f(x) u^i(x) dx \right), \\ \text{where} \quad & u^i \in H^1(\Omega_i),\ u^i = 0 \quad \text{on} \quad \Gamma_u^i,\ i = 1, \cdots, N, \\ & u^i - u^j \leq 0 \quad \text{on} \quad \partial\Omega_i \cap \partial\Omega_j, \forall\, i < j, \end{aligned} \tag{1}$$

with variable coefficients and multiple bodies $\Omega_i \subset \mathbb{R}^2$ with their boundaries and the Dirichlet boundaries denoted by $\partial\Omega_i$ and Γ_u^i, respectively, for $i = 1, \cdots, N$. The bodies are decomposed into subdomains,

$$\Omega_i = \bigcup_{j=1}^{N_i} \Omega_{i,j}, \quad i = 1, \cdots, N.$$

Here, *bodies* mean separate physical entities; for instance, two rubber balls in contact with each other are considered two bodies. *Subdomains*, on the other hand, is artificially introduced for convenience; a rubber ball can consist of as many subdomains

R. Bank et al. (eds.), *Domain Decomposition Methods in Science and Engineering XX*, Lecture Notes in Computational Science and Engineering 91,
DOI 10.1007/978-3-642-35275-1_35,

as the modeler wants. We assume that the coefficient ρ varies moderately within each body, $\Omega_i, i = 1, \cdots, N$. The diameters of Ω_i and $\Omega_{i,j}$ are denoted by H_i and $H_{i,j}$, respectively. The smallest diameters of any element in Ω_i and $\Omega_{i,j}$ are denoted by h_i and $h_{i,j}$, respectively. Also, $H_b := \max_i H_i$, $H_s := \max_{i,j} H_{i,j}$, $\frac{H_b}{h} := \max_i \frac{H_i}{h_i}$, $\frac{H_s}{h} := \max_{i,j} \frac{H_{i,j}}{h_{i,j}}$. We introduce the following:

$$\begin{aligned} \Gamma_{gl} &:= \bigcup_{i \neq j} \partial\Omega_i \cap \partial\Omega_j, \text{ potential contact surface between bodies,,} \\ \Gamma_{loc}^{(i)} &:= \bigcup_{j \neq k} (\partial\Omega_{i,j} \cap \partial\Omega_{i,k}), \text{interface between subdomains, } i = 1, \cdots, N. \end{aligned} \tag{2}$$

Here, the subscripts *gl* and *loc* stand for global and local, respectively, referring to nature of the interfaces. For each body, $\Omega_i, i = 1, \cdots, N$, two kinds of finite element spaces are introduced: $\widehat{W}^{(i)}$ is a standard finite element space of continuous, piecewise linear functions and, as such, is continuous across $\Gamma_{loc}^{(i)}$; $\widetilde{W}^{(i)}$ is a more general space, consisting of finite element functions required to be continuous only at the *primal* nodes (i.e., the vertex nodes of $\Gamma_{loc}^{(i)}$ in this two-dimensional case; more sophisticated continuity couplings, i.e., primal constraints, are required in $\widetilde{W}^{(i)}$ for three-dimensional problems; see [9, 10]), as in the FETI-DP (dual-primal FETI) method. The trace spaces of $\widetilde{W}^{(i)}$ and $\widehat{W}^{(i)}$ on $\Gamma_{loc}^{(i)} \cup (\partial\Omega_i \cap \Gamma_{gl})$ are denoted by $\widetilde{V}^{(i)}$ and $\widehat{V}^{(i)}$, respectively. The trace space of $\widehat{W}^{(i)}$ on $\partial\Omega_i \cap \Gamma_{gl}$ is denoted by $V_{OL}^{(i)}$, where OL stands for "one level." The Schur complements of the stiffness matrices for $\widetilde{W}^{(i)}$ and $\widehat{W}^{(i)}$, obtained by eliminating unknowns corresponding to the *subdomain interiors*, that is, those *not* associated with $\Gamma_{loc}^{(i)} \cup (\partial\Omega_i \cap \Gamma_{gl})$, are denoted by $\widetilde{S}_\Gamma^{(i)}$ and $\widehat{S}_\Gamma^{(i)}$, respectively. The Schur complement $S_{OL}^{(i)}$ of the stiffness matrix for $\widehat{W}^{(i)}$, on the other hand, is obtained by eliminating unknowns corresponding to the *body interior*, i.e., those *not* associated with $\partial\Omega_i \cap \Gamma_{gl}$. Therefore $\widetilde{S}_\Gamma^{(i)}, \widehat{S}_\Gamma^{(i)}$, and $S_{OL}^{(i)}$ can be viewed as operators on $\widetilde{V}^{(i)}, \widehat{V}^{(i)}$, and $V_{OL}^{(i)}$, respectively. We note that applying $S_{OL}^{(i)}$ requires solving a Dirichlet problem on Ω_i.

Let $\widetilde{V} := \prod_{i=1}^N \widetilde{V}^{(i)}, \widehat{V} := \prod_{i=1}^N \widehat{V}^{(i)}, V_{OL} := \prod_{i=1}^N V_{OL}^{(i)}, \widetilde{S} = \text{diag}_{i=1}^N \widetilde{S}_\Gamma^{(i)}, \widehat{S} = \text{diag}_{i=1}^N \widehat{S}_\Gamma^{(i)}$, and $S_{OL} := \text{diag}_{i=1}^N S_{OL}^{(i)}$. We also introduce matrices $\widetilde{B}, \widehat{B}$, and B_{OL}, with elements of $\{0, -1, 1\}$: $\widetilde{B}u \Leftrightarrow u \in \widetilde{V}$ is continuous across $\Gamma_{loc}^{(i)}, \forall i$, as well as Γ_{gl}; $\widehat{B}v \Leftrightarrow v \in \widehat{V}$ is continuous across Γ_{gl}; $B_{OL}w \Leftrightarrow w \in V_{OL}$ is continuous across Γ_{gl}.

2 Algorithms

With the matrices defined in Sect. 1, we can consider the following algorithm for solving (1):

Algorithm: Active set method + Krylov subspace method

1. Initialize u^0. Set $k = 0$. Set $\mathscr{A}_k$, a subset of the index set $\{1, \cdots, \#(\text{rows}(\widetilde{B}))\}$ (resp. $\#(\text{rows}(\widehat{B}))$), according to the active set method of choice.
2. Solve

$$\min_{u \in \widetilde{V}} \frac{1}{2} u^T \widetilde{S} u - \widetilde{g}^T u, \quad \text{with} \quad Z^k \widetilde{B} u = 0 \tag{3}$$

$$\left(\text{resp.} \min_{u \in \widehat{V}} \frac{1}{2} u^T \widehat{S} u - \widehat{g}^T u, \quad \text{with} \quad \widehat{Z}^k \widehat{B} u = 0\right) \tag{4}$$

   approximately to a given precision, using a Krylov subspace method. Set u^{k+1} to the resulting approximate solution. Find $\mathscr{A}_{k+1}$ accordingly.
3. Set $k = k + 1$. Stop if $\mathscr{A}_{k-1} = \mathscr{A}_k$; return to Step 2 otherwise.

Note that the linear problem in the kth iteration of the active set method is formulated as a minimization problem in terms of the interface variables in $\widetilde{V}$ or $\widehat{V}$. Here, $\widetilde{g} \in \widetilde{V}$ and $\widehat{g} \in \widehat{V}$ are appropriate load vectors. The square, diagonal matrix Z^k, with all elements equal to 0 or 1, is chosen such that $Z^k \widetilde{B} = \widetilde{B}_{\mathscr{A}_k}$, where $\widetilde{B}_{\mathscr{A}_k}$ is obtained by replacing the ith row of $\widetilde{B}$ with zeros for $\forall i \notin \mathscr{A}_k$. The matrix $\widehat{Z}^k$ is defined analogously. The minimization problems (3) and (4) are equivalent to the following saddle point problems,

$$\begin{bmatrix} \widetilde{S} & (Z^k \widetilde{B})^T \\ Z^k \widetilde{B} & 0 \end{bmatrix} \begin{bmatrix} u \\ \lambda \end{bmatrix} = \begin{bmatrix} \widetilde{g} \\ 0 \end{bmatrix}, \tag{5}$$

and

$$\begin{bmatrix} \widehat{S} & (\widehat{Z}^k \widehat{B})^T \\ \widehat{Z}^k \widehat{B} & 0 \end{bmatrix} \begin{bmatrix} u \\ \lambda \end{bmatrix} = \begin{bmatrix} \widehat{g} \\ 0 \end{bmatrix}, \tag{6}$$

respectively. We now consider the preconditioning of (5) and (6).

The **FETI-FETI** method is a combination of the one-level FETI method with a Dirichlet preconditioner [4] and the FETI-DP method [5], and was used in [1, 2] to solve frictionless contact problems. For (6), it is natural to follow the approach in the one-level and FETI-DP methods and form a Schur complement equation

$$\underbrace{Z^k \widetilde{B} \widetilde{S}^\dagger \widetilde{B}^T Z^k}_{:=F} \lambda = Z^k \widetilde{B} \widetilde{S}^\dagger \widetilde{g} + Z^k \widetilde{B} R \alpha, \tag{7}$$

where $\widetilde{S}^\dagger$ is a pseudoinverse of $\widetilde{S}$, $\text{range}(R) = \text{null}(\widetilde{S})$, and the vector α is to be determined. We solve (7) with the preconditioned conjugate gradient (PCG) method, using the following preconditioner:

$$P_F^{-1} := Z^k \widetilde{B}_D \widetilde{S} \widetilde{B}_D^T Z^k. \tag{8}$$

If $\widetilde{S}$ is singular, then the PCG method needs to be confined to the following subspace:

$$V^k := \{\lambda : Z^k \widetilde{B} \lambda \in \text{range}(\widetilde{S})\}. \tag{9}$$

Most of the computational work in each iteration of the PCG method goes into the applications of $\widetilde{S}^\dagger$ and $\widetilde{S}$, in the applications of F and P_F^{-1}, respectively. The application

of $\widetilde{S}$ involves solving a Dirichlet problem on each subdomain, $\Omega_{i,j}, i = 1, \cdots, N, j = 1, \cdots, N_i$. The application of $\widetilde{S}^{\dagger}$ involves solving a Dirichlet problem in each subdomain, with the Dirichlet boundary condition imposed only at subdomain vertices, plus solving a coarse problem on each body, associated with the set of vertices of $\Gamma_{loc}^{(i)}, i = 1, \cdots, N$; for details, see, e.g., [13],[14, Chap. 6].

The **hybrid** method is a combination of the one-level FETI method with a Dirichlet preconditioner and the BDDC (balancing domain decomposition by constraints) method [3]. For (6), forming a Schur complement equation similar to (7) is much more expensive because of the dense structure of $\widehat{S}$. Hence we keep the saddle point formulation (6) as is and solve it with the preconditioned conjugate residual (PCR) method. As in the FETI-FETI method, the PCR method needs to be confined to the following subspace:

$$\widehat{V}^k := \{\lambda : \widehat{Z}^k \widehat{B} \lambda \in \text{range}(\widehat{S})\}.$$

Letting P^k denote an orthogonal projection onto V^k, we rewrite (6) as

$$\underbrace{\begin{bmatrix} \widehat{S} & (P^k \widehat{Z}^k \widehat{B})^T \\ P^k \widehat{Z}^k \widehat{B} & 0 \end{bmatrix}}_{:=\mathscr{A}} \begin{bmatrix} u \\ \mu \end{bmatrix} = \begin{bmatrix} \widehat{g} - \widehat{B}^T \lambda_0 \\ 0 \end{bmatrix}, \tag{10}$$

with λ_0 satisfying $(\widehat{Z}^k \widehat{B}^T)\lambda_0 \in \text{range}(\widehat{S})$. For details on how to recover a solution of (6) from a solution of (10), see [8]. Letting P_R denote an orthogonal projection onto $\text{range}(\widehat{S})$, we introduce the preconditioner $\mathscr{B}$, where

$$\mathscr{B}^{-1} = \begin{bmatrix} P_R M_{BDDC}^{-1} P_R & 0 \\ 0 & P^k M_D^{-1} P^k \end{bmatrix}. \tag{11}$$

Here, M_{BDDC} is a block diagonal matrix consisting of the *BDDC* preconditioners [3] for the bodies:

$$M_{BDDC}^{-1} = \text{diag}_{i=1}^{N} M_{BDDC}^{(i)^{-1}} = \text{diag}_{i=1}^{N} \widetilde{R}_{D,\Gamma}^{(i)^T} \widetilde{S}_{\Gamma}^{(i)^{\dagger}} \widetilde{R}_{D,\Gamma}^{(i)},$$

where $\widetilde{R}_{D,\Gamma}^{(i)^T}, i = 1, \cdots, N$, is a scaled restriction from $\widetilde{V}^{(i)}$ to $\widehat{V}^{(i)}$, with the scaling factors determined by the material coefficients; similarly, $B_{OL,D}$ is a scaled version of B_{OL}. For details on the definition of these matrices, see, for instance, [11, 13]. Then M_D can be viewed as a Dirichlet preconditioner of the one-level FETI method, obtained by viewing each body, Ω_i, as a subdomain:

$$M_D^{-1} = \widehat{Z}^k B_{OL,D} S_{OL} B_{OL,D}^T \widehat{Z}^{k^T}.$$

Most of the computational work in each iteration of the PCR method goes into the application of $\widehat{S}$, in the application of $\mathscr{A}$, and the application of $\widetilde{S}_{\Gamma}^{(i)^{\dagger}}, i = 1, \cdots, N$ and S_{OL}, in the application of $\mathscr{B}^{-1}$. The application of $\widehat{S}$ requires solving a Dirichlet problem on each subdomain, $\Omega_{i,j}, i = 1, \cdots, N, j = 1, \cdots, N_i$. The application of

$\widetilde{S}_{\Gamma}^{(i)^{\dagger}}, i = 1, \cdots, N$, which is carried out in the FETI-FETI method as well, requires solving a Dirichlet problem on $\Omega_{i,j}, j = 1, \cdots, N_i$ with the Dirichlet boundary condition imposed only at the vertices, plus solving a coarse problem on Ω_i associated with the vertices of $\Gamma_{loc}^{(i)}$. The application of S_{OL}, however, requires solving a Dirichlet problem on each body, which is expensive; therefore in practice such a Dirichlet problem needs only to be solved inexactly, for instance with a Krylov subspace method. A preconditioner for solving such a Dirichlet problem is proposed and tested in [11].

3 Theory

We now present condition number estimates for the FETI-FETI and hybrid methods. Because of space limitations, details and proofs are given elsewhere; see [11, 12].

Theorem 1. *Let F, P_F, and V^k be defined as in (7) and (9), respectively. For any $\lambda \in V^k$, we have*

$$\langle P_F\lambda, \lambda\rangle \leq \langle F\lambda, \lambda\rangle \leq C(H_b/H_s)(1+\log(H_s/h))^2\langle P_F\lambda, \lambda\rangle,$$

where $C > 0$ is a constant independent of the sizes of the bodies, subdomains, and elements.

Convergence of the PCR method for the hybrid method is determined by

$$\mathscr{K}(\mathscr{B}^{-1}\mathscr{A}) := \frac{\mu_{max}}{\mu_{min}} = \frac{\max\{|\lambda| : \lambda \in \sigma(\mathscr{B}^{-1}\mathscr{A})\}}{\min\{|\lambda| : \lambda \in \sigma(\mathscr{B}^{-1}\mathscr{A})\}}, \tag{12}$$

where $\sigma(\mathscr{B}^{-1}\mathscr{A})$ is the spectrum of $\mathscr{B}^{-1}\mathscr{A}$ on range$(P_R) \times \widehat{V}^k$.

Theorem 2. *Let $\mathscr{B}^{-1}, \mathscr{A}$, and $\mathscr{K}(\mathscr{B}^{-1}\mathscr{A})$ be defined as in (11)–(12), respectively. We then have the following bound:*

$$\mathscr{K}(\mathscr{B}^{-1}\mathscr{A}) \leq C(1+\log(H_b/h))^2(1+\log(H_s/h))^2,$$

where $C > 0$ is a constant independent of the sizes of the bodies, subdomains, and elements.

4 Numerical Results: Auxiliary Linear Problems

We solve the following equality-constrained minimization problem:

$$\begin{aligned} &\min \sum_{i=1}^{N_b \times N_b} \left(\frac{1}{2}\int_{\Omega_i} |\nabla u^i|^2 dx - \int_{\Omega_i} f u^i dx \right), \\ &\text{with equality constraints to be specified,} \end{aligned} \tag{13}$$

Table 1. Results of FETI-FETI and hybrid.

			FETI-FETI				Hybrid	
			I		II		I	II
$1/H_b$	H_b/H_s	H_s/h	cond	iter	cond	iter	iter	iter
2	fixed	fixed	2.89	7	2.31	7	10	10
4	at 2	at 2	4.41	12	2.85	10	11	8
6			4.51	13	2.91	10	11	9
8			4.55	14	2.93	10	11	8
10			4.56	14	2.94	10	11	8
12			4.57	13	2.95	10	11	7
14			4.58	14	2.96	10	11	7
16			4.58	14	2.96	10	11	7
fixed	4	fixed	7.68	10	5.02	9	10	10
at 2	6	at 2	12.70	12	7.46	10	10	10
	8		17.80	13	8.12	10	10	10
	10		22.93	15	10.96	11	10	8
	12		28.08	16	13.43	12	10	8
	14		33.25	17	14.01	12	9	8
	16		38.41	17	16.90	12	8	7
fixed	fixed	4	4.71	9	4.73	9	12	11
at 2	at 2	6	5.90	10	6.37	10	13	13
		8	6.90	10	7.08	10	13	13
		10	7.79	11	8.27	11	14	14
		12	8.55	11	9.25	11	14	14
		14	9.23	12	9.71	12	14	14
		16	9.83	12	10.52	12	14	14

where $\Omega_i \subset \mathbb{R}^2, i = 1, \cdots, N_b \times N_b$ are square bodies with side length $H_b := 1/N_b$, which collectively form the domain $\bar{\Omega} = \bigcup_{i=1}^{N_b \times N_b} \bar{\Omega}_i = [0,1] \times [0,1]$. We require $u^i \in H^1(\Omega_i), u^i|_{\partial\Omega_i \cap \partial\Omega} = 0$. Each Ω_i is decomposed into $N_s \times N_s$ square subdomains, each of which is discretized by square bilinear elements of side length h. Also, $\Gamma := \cup_{i \neq j} \partial\Omega_i \cap \partial\Omega_j$ denotes the interface between the bodies.

We supplement (13) with two different equality constraints, associated with different *contact areas* between the bodies. In the first problem, the entire Γ is considered as the contact area, that is, we require the continuity of the displacement vector across the entire Γ. This case has already been considered by Klawonn and Rheinbach [6] and Klawonn and Rheinbach [7]. In the second problem, continuity is imposed only on the middle third of the faces between the bodies. We solve these problems with both the FETI-FETI and hybrid methods. The PCG and PCR iterations are stopped when the norm of the residual has been reduced by a factor of 10^{-6}.

The results are shown in Table 1. We have three parameters to vary: the number of bodies across Ω ($N_b = 1/H_b$), the number of subdomains across each body

($N_s = H_b/H_s$), and the number of elements across each subdomain (H_s/h). We vary one parameter while keeping the other two fixed. The results for the first set of experiments, with the entire Γ as the contact surface, are shown in column I; those for the second set of experiments with a reduced contact area are shown in column II.

Note the linear dependence of the condition number on the number of subdomains across each body, H_b/H_s, for the FETI-FETI method, which confirms our theoretical finding. Note also that the iteration counts of the hybrid method do not increase as the number of subdomains is increased. Similar numerical results for the FETI-FETI method have been obtained independently by Klawonn and Rheinbach [6] and Klawonn and Rheinbach [7].

Acknowledgments This work was supported by the Office of Advanced Scientific Computing Research, Office of Science, U.S. Department of Energy, under Contract DE-AC02-06CH11357.

Bibliography

[1] Philip Avery and Charbel Farhat. The FETI family of domain decomposition methods for inequality-constrained quadratic programming: Application to contact problems with conforming and nonconforming interfaces. *Computer Methods in Applied Mechanics and Engineering*, 198(21-26):1673–1683, 2009. Advances in Simulation-Based Engineering Sciences - Honoring J. Tinsley Oden.

[2] Philip Avery, Gert Rebel, Michel Lesoinne, and Charbel Farhat. A numerically scalable dual-primal substructuring method for the solution of contact problems–part I: the frictionless case. *Comput. Methods Appl. Mech. Engrg.*, 193(23-26):2403–2426, 2004.

[3] Clark R. Dohrmann. A preconditioner for substructuring based on constrained energy minimization. *SIAM J. Sci. Comput.*, 25(1):246–258, 2003.

[4] Charbel Farhat, Jan Mandel, and François-Xavier Roux. Optimal convergence properties of the FETI domain decomposition method. *Comput. Methods Appl. Mech. Engrg.*, 115(3-4):365–385, 1994.

[5] Charbel Farhat, Michel Lesoinne, Patrick LeTallec, Kendall Pierson, and Daniel Rixen. FETI-DP: a dual-primal unified FETI method. I. A faster alternative to the two-level FETI method. *Internat. J. Numer. Methods Engrg.*, 50(7):1523–1544, 2001.

[6] Axel Klawonn and Oliver Rheinbach. A hybrid approach to 3-level FETI. *PAMM Proc. Appl. Math. Mech.*, 8(1):10841–10843, 2008.

[7] Axel Klawonn and Oliver Rheinbach. Highly scalable parallel domain decomposition methods with an application to biomechanics. *ZAMM Z. Angew. Math. Mech.*, 90(1):5–32, 2010.

[8] Axel Klawonn and Olof B. Widlund. A domain decomposition method with Lagrange multipliers and inexact solvers for linear elasticity. *SIAM J. Sci. Comput.*, 22(4):1199–1219, 2000.

[9] Axel Klawonn and Olof B. Widlund. Dual-primal FETI methods for linear elasticity. *Comm. Pure Appl. Math.*, 59(11):1523–1572, 2006.

[10] Axel Klawonn, Olof B. Widlund, and Maksymilian Dryja. Dual-primal FETI methods for three-dimensional elliptic problems with heterogeneous coefficients. *SIAM J. Numer. Anal.*, 40(1):159–179, 2002.

[11] Jungho Lee. Two domain decomposition methods for auxiliary linear problems arising from an active-set based treatment of a multibody elliptic variational inequality. 2010. Accepted with minor revisions, *SIAM J. Sci. Comput.*

[12] Jungho Lee. *A Hybrid Domain Decomposition Method and its Applications to Contact Problems*. PhD thesis, Courant Institute of Mathematical Sciences, September 2009.

[13] Jing Li and Olof B. Widlund. FETI-DP, BDDC, and block Cholesky methods. *Internat. J. Numer. Methods Engrg.*, 66(2):250–271, 2006.

[14] Andrea Toselli and Olof Widlund. *Domain decomposition methods—algorithms and theory*, volume 34 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin, 2005.

New Theoretical Coefficient Robustness Results for FETI-DP

Clemens Pechstein[1], Marcus Sarkis[2], and Robert Scheichl[3]

[1] Institute of Computational Mathematics, Johannes Kepler University, Altenberger Str. 69, 4040 Linz, Austria, clemens.pechstein@numa.uni-linz.ac.at
[2] Mathematical Sciences Department, Worcester Polytechnic Institute, 100 Institute Road, Worcester, MA 01609-2280, United States, and Instituto de Matemática Pura e Aplicada (IMPA), Brazil, msarkis@wpi.edu
[3] Department of Mathematical Sciences, University of Bath, Bath BA2 7AY, United Kingdom, r.scheichl@maths.bath.ac.uk

1 Introduction

In this short note, we present new weighted Poincaré inequalities (WPIs) with weighted averages that allow a robustness analysis of dual-primal finite element tearing and interconnecting (FETI-DP) methods in certain cases where jumps of coefficients are not aligned with the subdomain partition.

Let Ω be a bounded Lipschitz domain in $\mathbb{R}^2$ or $\mathbb{R}^3$. We consider the weak form of the scalar elliptic PDE

$$-\mathrm{div}(\alpha \nabla u) \;=\; f \qquad \text{in } \Omega, \tag{1}$$

with a uniformly positive diffusion coefficient $\alpha \in L^\infty(\Omega)$ that is piecewise constant with respect to a (possibly rather fine) partitioning of Ω. The discretization by continuous and piecewise linear finite elements (FEs) on a mesh $\mathscr{T}(\Omega)$ leads to the sparse (but in general large) linear system

$$\mathbf{K}\mathbf{u} \;=\; \mathbf{f}.$$

We consider FETI-DP solvers (see [2, 4, 5]) for the fast (and parallel) solution of this system, and we follow the structure described in [12, Sect. 6.4]. To this end, we partition the domain Ω into non-overlapping subdomains Ω_i, $i = 1,\dots,N$ such that the global mesh $\mathscr{T}(\Omega)$ resolves the interface $\bigcup_{i\neq j} \partial\Omega_i \cap \partial\Omega_j$. The interface itself can be divided into subdomain vertices, edges, and faces (for $d = 3$), cf. [12, Sect. 4.2].

Without loss of generality, we assume that α is constant on each element of $\mathscr{T}(\Omega)$. Crucially, we do *not* assume that α is constant on each subdomain. However, we need assumptions on the *kind of jumps*. Let α_i denote the restriction of α to Ω_i and note that it has a well-defined trace in $L^2(\partial\Omega_i)$. For each subdomain edge (face)

R. Bank et al. (eds.), *Domain Decomposition Methods in Science and Engineering XX*, 313
Lecture Notes in Computational Science and Engineering 91,
DOI 10.1007/978-3-642-35275-1_36,

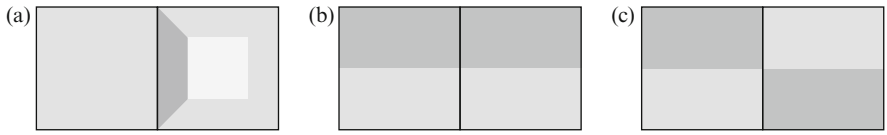


Fig. 1. Different types of coefficient jumps along an edge between two subdomains: (**a**) across (**b**) along (**c**) both across and along

$\mathscr{E}$ on Ω_i, let $V^h(\mathscr{E})$ denote the restriction of the global FE space to $\overline{\mathscr{E}}$ and let us define the weighted average

$$\overline{v}^{\mathscr{E},\alpha_i} := \frac{\int_{\mathscr{E}} \alpha_i v}{\int_{\mathscr{E}} \alpha_i} \qquad \text{for } v \in V^h(\mathscr{E}). \tag{2}$$

Assumption A1. Whenever two Ω_i and Ω_j share an edge (face) $\mathscr{E}$, the weighted averages of any function $v \in V^h(\mathscr{E})$ coincide: $\overline{v}^{\mathscr{E},\alpha_i} = \overline{v}^{\mathscr{E},\alpha_j}$.

A sufficient condition for Assumption A1 is that the coefficient jumps *either across or along*, but not both at the same time. For an illustration see Fig. 1. Our assumptions rules out situations of type (c).

Following [12, Algorithm B], we define the *primal space* $\widehat{W}_\Pi$ spanned by the vertex nodal basis functions at subdomain vertices, the subdomain edge cut-off functions and subdomain face cut-off functions (all of them extended discrete α-harmonically from the interface to the subdomain interiors). The *dual space* W_Δ contains FE functions that are discontinuous across the subdomain interfaces with vanishing α-weighted averages over the subdomain faces, edges, and vertices. We formally perform a change of basis, such that we have a splitting of the degrees of freedom (DOFs) into primal and dual ones, and work in the space $\widetilde{W} = \widehat{W}_\Pi \oplus W_\Delta$.

Let $B : \widetilde{W} \to U$ be the usual jump operator. The FETI-DP system

$$F\lambda \;=\; B\widehat{K}^{-1}\widehat{f} \tag{3}$$

is solved by preconditinioned conjugate gradients, where $F := B\widehat{K}^{-1}B^\top$ and where $\widehat{K}$, $\widehat{f}$ denote the stiffness matrix and load vector partially assembled at the primal DOFs, respectively. The overall solution is then given by

$$u \;=\; \widehat{K}^{-1}(\widehat{f} - B^\top\lambda).$$

Next, we define a FETI-DP preconditioner that is slightly modified to allow for certain coefficient jumps (cf. [3, 7]). Let $i = 1,\dots,N$ be fixed and let $\mathscr{T}(\Omega_i)$ denote the mesh restricted to subdomain Ω_i. For each mesh node x^h on $\overline{\Omega}_i$, we set

$$\widehat{\alpha}_i(x^h) := \max_{T\in\mathscr{T}(\Omega_i):x^h\in\overline{T}} \alpha_{|T}\,. \tag{4}$$

Furthermore, if $\mathscr{N}_{x^h}$ denotes the index set of subdomains sharing the mesh node x^h, we define the weighted counting function

$$\delta_i^\dagger(x^h) := \begin{cases} \dfrac{\widehat{\alpha}_i(x^h)}{\sum_{j\in\mathcal{N}_{x^h}} \widehat{\alpha}_j(x^h)}, & \text{if } x^h \text{ lies on } \overline{\Omega}_i\,, \\ 0, & \text{otherwise.} \end{cases}$$

Using these counting functions we define the scaled jump operator B_D according to [12, Sect. 6.4.1] (for details see also [9] where the same scaled jump operator was used to define a one-level FETI preconditioner). The FETI-DP preconditioner is finally given by

$$M^{-1} := B_D S B_D^\top, \tag{5}$$

where $S = \mathrm{diag}(S_i)_{i=1}^N$ is the block-diagonal Schur complement of the block stiffness matrix $K = \mathrm{diag}(K_i)_{i=1}^N$, eliminating the interior DOFs in each subdomain. Alternatively, one may replace B and B_D in (3), (5) by the respective operators which only act on the dual DOFs, which reduces the number of redundancies in λ.

2 Weighted Poincaré Inequalities with Weighted Averages

Let D be a bounded Lipschitz polytope and let $\{Y_\ell\}_{\ell=1}^n$ be a subdivision of D into open Lipschitz polytopes such that

$$\alpha_{|Y_\ell} \;=\; \alpha_\ell \;=\; \text{const.} \tag{6}$$

Furthermore, let $\mathscr{X} \subset \partial D$ be a manifold of dimension $0 \le d_{\mathscr{X}} \le d-1$ (usually a vertex, an open subdomain edge or an open face, or a union of these). We define

$$\mathscr{X}_\ell := \overline{Y}_\ell \cap \mathscr{X}\,.$$

Some of these sets may be empty or have lower dimension than $\mathscr{X}$. However, with the index set $I_{\mathscr{X}} := \{\ell : \mathrm{meas}_{d_{\mathscr{X}}}(\mathscr{X}_\ell) > 0\}$ we can write

$$\overline{\mathscr{X}} \;=\; \bigcup\nolimits_{k\in I_{\mathscr{X}}} \overline{\mathscr{X}}_k\,.$$

In general, for different indices k, $\ell \in I_{\mathscr{X}}$, the manifolds $\mathscr{X}_k$ and $\mathscr{X}_\ell$ may have a non-trivial intersection or even coincide. For simplicity, we assume that

$$k \neq \ell \in I_{\mathscr{X}} \quad\Longrightarrow\quad \mathrm{meas}_{d_{\mathscr{X}}}(\mathscr{X}_k \cap \mathscr{X}_\ell) \;=\; 0.$$

The general case needs more formalism and will be treated in an upcoming paper [10]. Finally, we can define a meaningful trace $\alpha_{\mathrm{tr}} \in L^\infty(\mathscr{X})$ of α by

$$\alpha_{\mathrm{tr}}(x) \;=\; \alpha_k \qquad \text{for } x \in \mathscr{X}_k\,.$$

Let $\{V^h(D)\}_h$ be a family of H^1-conforming FE spaces associated with a quasi-uniform family of triangulations of D. For $v \in V^h(D)$, we define the weighted (semi)norms and the weighted average on $\mathscr{X}$ by

$$\|v\|^2_{L^2(D),\alpha} := \int_D \alpha\, v^2\,, \quad |v|^2_{H^1(D),\alpha} := \int_D \alpha\, |\nabla v|^2 \quad \text{and} \quad \overline{v}^{\mathscr{X},\alpha_{\mathrm{tr}}} := \frac{\int_{\mathscr{X}} \alpha_{\mathrm{tr}}\, v}{\int_{\mathscr{X}} \alpha_{\mathrm{tr}}}\,.$$

We are interested in the following WPI with weighted average:

$$\|u - \overline{u}^{\mathscr{X},\alpha_{\mathrm{tr}}}\|^2_{L^2(D),\alpha} \;\le\; C_{P,\alpha}(D,\mathscr{X};h)\,\mathrm{diam}(D)^2\,|u|^2_{H^1(D),\alpha} \quad \forall u \in V^h(D). \tag{7}$$

In particular, we are interested under which assumptions the parameter $C_{P,\alpha}(D,\mathscr{X};h)$ is independent of the values $\{\alpha_\ell\}$.

Sufficient conditions for robustness. We need two crucial assumptions for (7) to be independent of the values $\{\alpha_\ell\}$. The first assumption is a quasi-monotonicity assumption on α. It has been introduced in [1] and generalized in [4, 8]. The second assumption states that $\mathscr{X}$ "sees" the largest coefficient.

Definition 1. *Let $0 \le m < d$ and let $\ell^* := \operatorname{argmax}_{1\le\ell\le s} \alpha_\ell$ denote the index of the largest coefficient.*[4]

(a) We call the region $P_{\ell_1,\ell_s} := (\overline{Y}_{\ell_1} \cup \ldots \cup \overline{Y}_{\ell_s})^\circ$, $1 \le \ell_1,\ldots,\ell_s \le n$ a type-m quasi-monotone path *from Y_{ℓ_1} to Y_{ℓ_s} (with respect to α), if*
  (i) the regions Y_{ℓ_i} and $Y_{\ell_{i+1}}$ share a common m-dimensional manifold, and
  (ii) $\alpha_{\ell_1} \le \alpha_{\ell_2} \le \ldots \le \alpha_{\ell_s}$.

(b) We say that α is type-m quasi-monotone *on D, if for all $k = 1,\ldots,n$ there exists a quasi-monotone type-m path from Y_k to Y_{ℓ^*}.*

Assumption A2. α is type-m quasi-monotone on D for some $0 \le m < d$.

Assumption A3. $\mathrm{meas}_{d_{\mathscr{X}}}(\mathscr{X} \cap \overline{Y}_{\ell^*}) > 0$.

In order to formulate our main theorem, we first need some definitions of generalized Poincaré constants/parameters.

Definition 2. *(i) For any bounded Lipschitz domain $Y \subset \mathbb{R}^d$ let $C_P(Y)$ be the smallest constant such that*

$$\|v - \overline{v}^Y\|^2_{L^2(Y)} \;\le\; C_P(Y)\,\mathrm{diam}(Y)^2\,|v|^2_{H^1(Y)} \qquad \forall v \in H^1(Y).$$

(ii) Let Z be the finite union of bounded Lipschitz polytopes such that $\overline{Z}$ is connected, and let $\{\mathscr{T}^h(Z)\}_h$ be a quasi-uniform family of triangulations of Z with the associated continuous piecewise linear FE spaces $\{V^h(Z)\}_h$. Let X, $W \subset \overline{Z}$ be manifolds/subdomains of (possibly different) dimension $\in \{0,\ldots,d\}$. Let $C_P(Z,X,W;h)$ be the best parameter such that

$$\|v - \overline{v}^X\|^2_{L^2(W)} \;\le\; C_P(Z,X,W;h)\,\frac{|W|}{|Z|}\,\mathrm{diam}(Z)^2\,|u|^2_{H^1(Z)} \qquad \forall v \in V^h(Z).$$

$|W|$ and $|Z|$ denote the measures of W and Z (in the respective dimension).

[4] We can assume without loss of generality that ℓ^* is unique. By definition, type-m quasi-monotonicity implies that otherwise all maximal subregions can be combined into a single subregion.

If Z is connected and if the dimensions of X and W are $\geq d-1$, we can define a constant $C_P(Z,X,W)$ independent of the discretization parameter h such that the inequality in Definition 2(ii) holds for all functions in $H^1(Z)$.

Theorem 1. *Let Assumptions A2 and A3 be satisfied. Then the parameter $C_{P,\alpha}(D,\mathscr{X};h)$ in formula (7) is independent of the values $\{\alpha_\ell\}_{\ell=1}^n$ and*

$$C_{P,\alpha}(D,\mathscr{X};h) \;\leq\; 2\left[C^{*,1}(h)+C^{*,2}(h)\right] \tag{8}$$

with

$$C^{*,1}(h) := \sum_{\ell=1}^{n} \frac{|Y_\ell|\,\mathrm{diam}(P_{\ell,\ell^*})^2}{|P_{\ell,\ell^*}|\,\mathrm{diam}(D)^2}\, C_P(P_{\ell,\ell^*},\mathscr{X}_{\ell^*},Y_\ell;h),$$

$$C^{*,2}(h) := \frac{|D|}{|\mathscr{X}_{\ell^*}|} \sum_{k\in I_{\mathscr{X}}} \frac{|\mathscr{X}_k|\,\mathrm{diam}(P_{k,\ell^*})^2}{|P_{k,\ell^*}|\,\mathrm{diam}(D)^2}\, C_P(P_{k,\ell^*},\mathscr{X}_{\ell^*},\mathscr{X}_k;h).$$

Proof. Without loss of generality, we may assume that $\overline{u}^{\mathscr{X},\alpha_{\mathrm{tr}}}=0$. For each index $\ell=1,\ldots,n$,

$$\tfrac{1}{2}\,\|u\|^2_{L^2(Y_\ell)} \;\leq\; \|u-\overline{u}^{\mathscr{X}_{\ell^*}}\|^2_{L^2(Y_\ell)} + |Y_\ell|\left(\overline{u}^{\mathscr{X}_{\ell^*}}\right)^2.$$

Due to Assumption A2, there is a quasi-monotone path from Y_ℓ to Y_{ℓ^*}. With $c_{\ell,\ell^*} := C_P(P_{\ell,\ell^*},\mathscr{X}_{\ell^*},Y_\ell;h)$, summation over $\ell=1,\ldots,n$ yields

$$\tfrac{1}{2}\,\|u\|^2_{L^2(D),\alpha} \;\leq\; \sum_{\ell=1}^{n} c_{\ell,\ell^*}\frac{|Y_\ell|}{|P_{\ell,\ell^*}|}\,\mathrm{diam}(P_{\ell,\ell^*})^2 \underbrace{\alpha_\ell |u|^2_{H^1(P_{\ell,\ell^*})}}_{\leq |u|^2_{H^1(D),\alpha}} + \underbrace{\sum_{\ell=1}^{n}\alpha_\ell\,|Y_\ell|}_{\leq \alpha_{\ell^*}|D|}\left(\overline{u}^{\mathscr{X}_{\ell^*}}\right)^2,$$

where we have used Definition 2(ii) and the quasi-monotonicity of P_{ℓ,ℓ^*}. The first sum is bounded by $C^{*,1}(h)\,\mathrm{diam}(D)^2\,|u|^2_{H^1(D),\alpha}$. To bound the remaining term, we use Cauchy's inequality and the definition of α_{tr}:

$$\alpha_{\ell^*}|D|\left(\overline{u}^{\mathscr{X}_{\ell^*}}\right)^2 \;\leq\; \frac{|D|}{|\mathscr{X}_{\ell^*}|}\,\alpha_{\ell^*}\|u\|^2_{L^2(\mathscr{X}_{\ell^*})} \;\leq\; \frac{|D|}{|\mathscr{X}_{\ell^*}|}\,\|u\|^2_{L^2(\mathscr{X}),\alpha_{\mathrm{tr}}}.$$

A variational argument yields

$$\begin{aligned}\|u\|^2_{L^2(\mathscr{X}),\alpha_{\mathrm{tr}}} \;&\leq\; \|u-\underbrace{\overline{u}^{\mathscr{X},\alpha_{\mathrm{tr}}}}_{=0}\|^2_{L^2(\mathscr{X}),\alpha_{\mathrm{tr}}} \;=\; \inf_{c\in\mathrm{R}}\|u-c\|^2_{L^2(\mathscr{X}),\alpha_{\mathrm{tr}}}\\ &\leq\; \|u-\overline{u}^{\mathscr{X}_{\ell^*}}\|^2_{L^2(\mathscr{X}),\alpha_{\mathrm{tr}}} \;=\; \sum_{k\in I_{\mathscr{X}}}\alpha_k\,\|u-\overline{u}^{\mathscr{X}_{\ell^*}}\|^2_{L^2(\mathscr{X}_k)}.\end{aligned}$$

Now, we have

$$\alpha_k\,\|u-\overline{u}^{\mathscr{X}_{\ell^*}}\|^2_{L^2(\mathscr{X}_k)} \;\leq\; C_P(P_{k,\ell^*},\mathscr{X}_{\ell^*},\mathscr{X}_k;h)\,\frac{|\mathscr{X}_k|}{|P_{k,\ell^*}|}\,\mathrm{diam}(P_{k,\ell^*})^2\,\alpha_k\,|u|^2_{H^1(P_{k,\ell^*})}.$$

Using the quasi-monotonicity of α on P_{k,ℓ^*} finally leads to (8).

Necessity of the conditions. As discussed in [8, Sect. 3.1], Assumption A2 is necessary to ensure that $C_{P,\alpha}(D,\mathscr{X};h)$ is independent of the values $\{\alpha_\ell\}$.

To see that A3 is necessary as well, assume that $\mathrm{meas}_{d_{\mathscr{X}}}(\mathscr{X}\cap\overline{Y}_{\ell^*})=0$. We choose a function u which is one on Y_{ℓ^*}. Since the average functional $v\mapsto \overline{v}^{\mathscr{X},\alpha_{\mathrm{tr}}}$ is independent of α_{ℓ^*}, we can prescribe values of u on $\mathscr{X}$ such that $\overline{u}^{\mathscr{X},\alpha_{\mathrm{tr}}}=0$ and continuously extend u into $D\subset\overline{Y}_{\ell^*}$. The whole construction of u is independent of α_{ℓ^*}, Since $\nabla u=0$ on Y_{ℓ^*}, the seminorm $|u|_{H^1(D),\alpha}$ is independent of α_{ℓ^*} as well. However, $\|u\|^2_{L^2(D),\alpha}\geq\alpha_{\ell^*}|Y_{\ell^*}|$. Therefore, if $\alpha\leq\alpha_k$ on $D\backslash Y_{\ell^*}$, then $C_{P,\alpha}(D,\mathscr{X};h)=\mathscr{O}\big(\frac{\alpha_{\ell^*}}{\alpha_k}\big)$ for $\alpha_{\ell^*}/\alpha_k\to\infty$. This means that Assumptions A2 and A3 in some sense *characterize* the robustness of the WPI with weighted average.

3 Robustness Proof of FETI-DP

To analyze the robustness of FETI-DP, we need the following assumption.

Assumption A4. For each subdomain Ω_i and for each subdomain edge (face) $\mathscr{E}$ of Ω_i, there is a Lipschitz domain $D_{i,\mathscr{E}}\subset\Omega_i$, such that $\mathscr{E}\subset\partial D_{i,\mathscr{E}}$ and Assumptions A2 and A3 are satisfied for $D=D_{i,\mathscr{E}}$ and $\mathscr{X}=\mathscr{E}$. The union of all the regions $D_{i,\mathscr{E}}$ covers a boundary layer Ω_{i,η_i} of width $\eta_i\geq h$ of Ω_i (see e.g. [6, Definition 2.6]).

Theorem 2. *Let Assumptions A1 and A4 hold. Then the condition number $\kappa(M^{-1}F)$ for the FETI-DP method is independent of the values of the coefficient α, in particular of any non-resolved jumps.*

Due to space limitations we only give a sketch of the proof. A detailed proof will be given in [10], together with a more detailed statement of Theorem 2 that makes precise the dependence of $\kappa(M^{-1}F)$ on geometric parameters, such as the ratios $\mathrm{diam}(\Omega_i)/h$ and $\mathrm{diam}(\Omega_i)/\eta_i$.

Let $\mathscr{H}_i$ denote the discrete α-harmonic extension from $\partial\Omega_i$ to Ω_i and let

$$|w|_S^2 := \sum_{i=1}^N |\mathscr{H}_i w|^2_{H^1(\Omega_i),\alpha}\,.$$

Then, following [12, Sect. 6.4.3], a bound of the kind

$$|P_D\,w|_S^2 \;\leq\; \omega\,|w|_S^2 \qquad \forall w\in\widetilde{W}, \tag{9}$$

where $P_D:=B_D^\top B$, implies that $\kappa(M^{-1}F)\leq\omega$.

As in the proof of [9, Lemma 5.6; formula (5.24)], we can introduce a set of cut-off functions associated with each subdomain edge (face) $\mathscr{E}$ whose support is contained in $D_{i,\mathscr{E}}$. It then follows that, for any $w\in\widehat{W}_\Pi\oplus W_\Delta$,

$$|P_D\,w|_S^2 \;\leq\; C\sum_{i=1}^N\Big[|\mathscr{H}_i w_i|^2_{H^1(\Omega_i),\alpha}+\sum_{\mathscr{E}}\frac{1}{\mathrm{diam}(\Omega_i)^2}\,\|\mathscr{H}_i w_i-\overline{w_i}^{\mathscr{E}}\|^2_{L^2(D_{i,\mathscr{E}}),\alpha}\Big],$$

where C depends on $\mathrm{diam}(\Omega_i)/h$ and $\mathrm{diam}(\Omega_i)/\eta_i$, but it is independent of the values $\{\alpha_\ell\}$. By Theorem 1, we can bound each of the weighted L^2 norms by the weighted H^1 seminorm of $\mathscr{H}_i w_i$, and thus obtain (9).

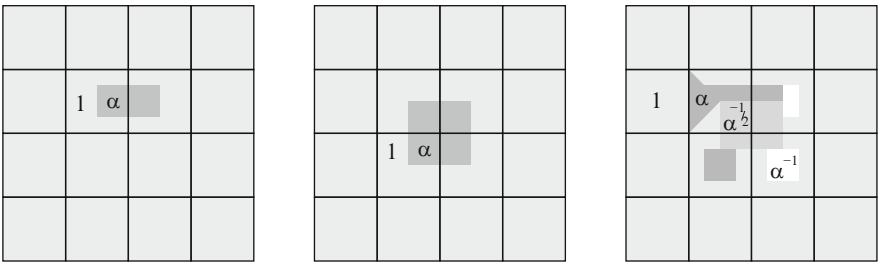


Fig. 2. Edge-island (*left*), cross-point island (*middle*), complicated coefficient (*right*)

α	condition	#iterations
1	1.58	10
10^1	1.57	10
10^3	1.56	10
10^5	1.56	10
10^7	1.56	10
10^{-1}	1.70	10
10^{-3}	1.74	10
10^{-5}	1.74	10
10^{-7}	1.74	11

α	condition	#iterations
1	1.58	10
10^1	1.59	10
10^3	1.59	10
10^5	1.59	10
10^7	1.59	10
10^{-1}	1.57	10
10^{-3}	1.57	10
10^{-5}	1.57	10
10^{-7}	1.57	10

α	condition	#iterations
1	1.58	10
10^1	1.61	11
10^2	1.62	11
10^3	1.62	11
10^4	1.62	11
10^{-1}	1.62	11
10^{-2}	1.60	11
10^{-3}	1.59	11
10^{-4}	1.59	11

Table 1. Edge-island (left), crosspoint-island (middle), complicated coefficient (right), $H/h = 32$.

4 Numerical Results

We provide results for the three examples shown in Fig. 2. Note that in the last example, the coefficient is not quasi-monotone on one of the subdomains, but satisfies Assumptions A1 and A4. In our implementation we used PARDISO [11]. The estimated condition numbers and the number of PCG iterations are displayed in Table 1. They clearly confirm Theorem 2.

5 Conclusion

We analyse a FETI-DP method for the scalar elliptic PDE (1) with possible jumps in the diffusion coefficient alpha. We show that provided weighted edge/face averages are used, the condition number of the preconditioned system is independent of coefficient jumps. The essential assumptions are A1 and A4, i.e., the coefficient does not jump both across and along any interfaces between two subdomains and the coefficient is quasi-monotone in the vicinity of any edge/face within each subdomain. The key theoretical tool that is of interest in itself is a novel weighted Poincaré inequality for functions with suitably chosen vanishing weighted face/edge averages. We are able to show that under Assumption A4, the Poincare constant of each neighborhood $D_{i,\mathscr{E}}$ can be bounded independent of jumps.

As in our previous work [8], the Poincaré constants (and thus also the condition number) will also depend on the "geometry" of the coefficient variation. In particular, for piecewise constant coefficients it will in general depend on the geometry of the subregions where the coefficient is constant. We did not give details of this dependence here, but this will be done in an upcoming paper [10] (using [8]). Cases where the coefficient jumps both along and across subdomain interfaces appear to be substantially harder to be treated and are also the subject of our future investigations.

Acknowledgments The authors would like to thank Clark Dohrmann for the fruitful discussions during and after the DD20 conference.

Bibliography

[1] M. Dryja, M. V. Sarkis, and O. B. Widlund. Multilevel Schwarz methods for elliptic problems with discontinuous coefficients in three dimensions. *Numer. Math.*, 72(3):313–348, 1996.

[2] C. Farhat, M. Lesoinne, P. Le Tallec, K. Pierson, and D. Rixen. FETI-DP: a dual-primal unified FETI method. I. A faster alternative to the two-level FETI method. *Internat. J. Numer. Methods Engrg.*, 50(7):1523–1544, 2001.

[3] A. Klawonn and O. Rheinbach. Robust FETI-DP methods for heterogeneous three dimensional elasticity problems. *Comput. Methods Appl. Mech. Engrg.*, 196(8):1400–1414, 2007.

[4] A. Klawonn, O. B. Widlund, and M. Dryja. Dual-primal FETI methods for three-dimensional elliptic problems with heterogeneous coefficients. *SIAM J. Numer. Anal.*, 40(1):159–179, 2002.

[5] J. Mandel and R. Tezaur. On the convergence of a dual-primal substructuring method. *Numer. Math.*, 88(3):543–558, 2001.

[6] C. Pechstein and R. Scheichl. Analysis of FETI methods for multiscale PDEs. *Numer. Math.*, 111(2):293–333, 2008.

[7] C. Pechstein and R. Scheichl. Scaling up through domain decomposition. *Appl. Anal.*, 88(10):1589–1608, 2009.

[8] C. Pechstein and R. Scheichl. Weighted Poincaré inequalities. *IMA J. Numer. Anal.*, published online 15 October 2012, DOI 10.1093/imanum/drs017.

[9] C. Pechstein and R. Scheichl. Analysis of FETI methods for multiscale PDEs – part II: interface variation. *Numer. Math.*, 118(1):485–529, 2011.

[10] C. Pechstein, M. Sarkis, and R. Scheichl. Analysis of FETI-DP methods for multiscale PDEs. In preparation, 2013.

[11] O. Schenk and K. Gärtner. On fast factorization pivoting methods for sparse symmetric indefinite systems. *Electron. Trans. Numer. Anal.* 23:158–179, 2006.

[12] A. Toselli and O. Widlund. *Domain Decomposition Methods – Algorithms and Theory*, volume 34 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin, 2005.

Monotone Multigrid Methods Based on Parametric Finite Elements

Thomas Dickopf and Rolf Krause

University of Lugano, Institute of Computational Science
Via G. Buffi 13, 6904 Lugano, Switzerland
thomas.dickopf@usi.ch, rolf.krause@usi.ch

Summary. In this paper, a particular technique for the application of elementary multilevel ideas to problems with warped boundaries is studied in the context of the numerical simulation of elastic contact problems. Combining a general multilevel setting with a different perspective, namely an advanced geometric modeling point of view, we present a (monotone) multigrid method based on a hierarchy of parametric finite element spaces. For the construction, a full-dimensional parameterization of high order is employed which accurately represents the computational domain.

The purpose of the volume parametric finite element discretization put forward here is two-fold. On the one hand, it allows for an elegant multilevel hierarchy to be used in preconditioners. On the other hand, it comes with particular advantages for the modeling of contact problems. After all, the long-term objective lies in an increased flexibility of *hp*-adaptive methods for contact problems.

1 Introduction

In the numerical simulation of elastic contact problems, the treatment of the non-penetration conditions at the potential contact boundary is of particular importance for both the quality of a finite element approximation and the overall efficiency of the algorithms. A vital challenge is to achieve an accurate description of geometric features, e.g., of warped surfaces, often incorporated in three-dimensional models from computer-aided design (CAD). Here, we investigate a new connection of different numerical methods, namely modern discretization techniques for partial differential equations on complex geometries on the one side and fast multilevel solvers for constrained minimization problems on the other side.

It is fair to say that the development of *hp*-adaptive methods for contact problems has not yet reached a mature state; see, e.g., [2] and the references therein. Partly, this is due to the difficulties concerning the geometric representation of the computational domain. A generally accepted paradigm is, though, that high order (finite element or boundary element) methods need high order meshes [11, 14]. This is especially difficult for three-dimensional multi-body contact problems. In this case, the application of non-conforming domain decomposition techniques [16] to realize

R. Bank et al. (eds.), *Domain Decomposition Methods in Science and Engineering XX*, Lecture Notes in Computational Science and Engineering 91,
DOI 10.1007/978-3-642-35275-1_37,

an optimal information transfer across geometrically non-matching warped contact interfaces is a highly demanding task. For low order finite elements, this has been achieved, among others, by the authors; see [6].

The perspective we offer here is a parametric finite element method. For *hp*-adaptive methods, it is convenient to have a parameterization describing the geometry accurately ready to hand. This is because a change of the computational domain due to locally altered polynomial degree is not desirable. Therefore, it is reasonable to uncouple the representation of the geometry on the one hand and of a scale of approximation spaces for the discrete solution on the other hand. These two purposes are usually not separated properly. But of course, one can find curved elements of other than isoparametric structure in some form or another in the literature; see, e.g., [8, 17] or the monograph [3] and the references therein. Note that, for similar reasons, an "isogeometric" concept, which uses NURBS bases for both the description of the geometry and the discrete solution of the differential equation, has been introduced in [11].

For practical computations, the development of fast and robust solvers is equally important. As this issue has not yet been in the main focus of, e.g., the isogeometric analysis [11], we would like to contribute ideas from the field of multilevel methods for variational inequalities. More precisely, we show how to use a monotone multigrid method to efficiently solve the non-linear contact problem discretized with low order parametric finite elements. Note that the actual treatment of higher order elements is beyond the scope of the present discussion.

To obtain multilevel parametric finite element spaces in case $d = 3$, we use a full-dimensional parameterization, constructed by tetrahedral transfinite interpolation [15] of CAD data, to lift standard Lagrange elements to the computational domain. Note that, similarly, a surface parameterization has been used in a wavelet Galerkin scheme for boundary integral equations; see [10]. Such a procedure may serve as an essential prerequisite to tackle the problems mentioned above. In particular, many of the issues arising in the generation of *p*-version meshes for curved boundaries [14] can be avoided in a quite elegant way. In this sense, although rather expensive, the use of a high order parameterization permits maximal freedom in an *hp*-adaptive discretization scheme. We presume that the present concept can also be combined with the ideas in [6].

All in all, our results constitute real progress made in the development of an efficient *hp*-adaptive simulation environment for elastic contact problems in case of complex three-dimensional geometries.

2 Parametric Finite Elements

In this section, we introduce a parametric finite element discretization. On the one hand, this method uses much more geometric information from a CAD model than standard finite elements; on the other hand, we do not use the same functions for the discrete approximation of the displacement field as for the representation of the geometry, which is done in the so-called "isogeometric analysis" introduced in [11]. We

use the associated space hierarchy in Sect. 3 to build a monotone multigrid method for low order elements.

In the following, the symbols φ with some indices stand for certain full-dimensional parameterizations or finite element transformations. We denote the (closed) d-simplex by Δ^d and its faces by Δ^d_j, $j \in \{1,\dots,d+1\}$. To describe the elastic body (here, $d = 3$) by a practicable parameterization, we consider a non-overlapping simplicial decomposition of the computational domain $\Omega \subset \mathbb{R}^d$ into a fixed number of $K \geq 1$ subdomains. Formally this reads as

$$\overline{\Omega} = \bigcup_{k=1}^{K} \overline{\Omega}_k = \bigcup_{k=1}^{K} \varphi_k(\Delta^d),$$

where the notation already indicates that the subdomains $(\Omega_k)_{k=1,\dots,K}$ appear as particular images of the simplex Δ^d under suitable parameterizations $(\varphi_k)_{k=1,\dots,K}$. This is illustrated in Fig. 1 (right).

Let us assume that the faces of the simplicial cells Ω_k, namely the surfaces $\varphi_k(\Delta^d_j)$, $k \in \{1,\dots,K\}$, $j \in \{1,\dots,d+1\}$, are given as B-patches. This way to represent polynomial surfaces is analyzed in [4]. In this case, the author of [15] proposes to construct the full-dimensional mappings $\varphi_k : \Delta^d \to \mathbb{R}^d$, $k \in \{1,\dots,K\}$, as transfinite interpolations of the surface values from the CAD model using certain blending functions. Particularly, the single parameterizations are smooth and they match across these B-patch surfaces if the surfaces themselves match. This gives rise to a consistent global parameterization which we do not write down explicitly. We note that this global mapping is continuous but not necessarily differentiable across the interior interfaces. In addition, one can guarantee that each parameterization φ_k satisfies the regularity assumption

$$\det(\nabla \varphi_k) > 0 \quad \text{in } \Delta^d. \tag{1}$$

In fact, this is one of the main results of [15].

In the following, we define the parametric finite element spaces in a rather straightforward way via a lift of standard Lagrange finite elements. For this purpose, let $(\mathscr{T}^k_\ell)_{\ell\in\mathbb{N}}$ be a family of nested simplicial meshes of Δ^d for each $k \in \{1,\dots,K\}$. To keep the global finite element spaces conforming, we assume that, at each level $\ell \in \mathbb{N}$, the meshes meeting at the faces of the simplicial subdomains Ω_k of Ω match. Let $\widehat{T}$ be the reference element; here, $\widehat{T} = \Delta^d$. Then, for each $T_\Delta \in \mathscr{T}^k_\ell$, there is an affine mapping $\varphi_{T_\Delta} : \widehat{T} \to \Delta^d$ such that $\varphi_{T_\Delta}(\widehat{T}) = T_\Delta$.

Now, we give a concise description of the parametric elements in Ω by employing the special finite element transformations

$$\varphi_T := \varphi_k \circ \varphi_{T_\Delta} : \widehat{T} \to \mathbb{R}^d, \tag{2}$$

which are diffeomorphisms between the reference element $\widehat{T}$ and the actual elements. That way, the parametric elements at level $\ell \in \mathbb{N}$ are identified as the images of the elements of the meshes $(\mathscr{T}^k_\ell)_{k=1,\dots,K}$; see Fig. 1. More precisely, a family of parametric meshes $(\mathscr{T}_\ell)_{\ell\in\mathbb{N}}$ of Ω can be defined by

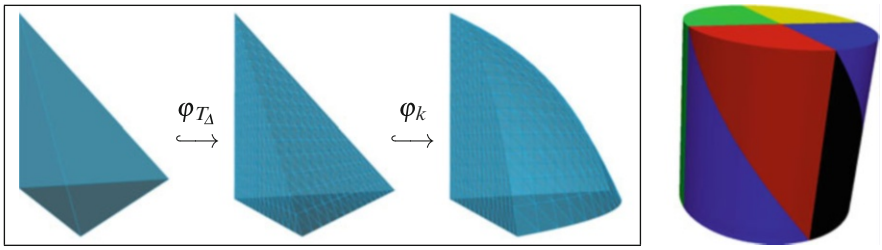


Fig. 1. From left to right: the reference element $\widehat{T} = \Delta^3$; a mesh of the simplex Δ^3; a parametric mesh (here, $K = 1$) where each element is an image of an affine element; a tetrahedral decomposition of a cylinder with $K = 8$

$$\mathscr{T}_\ell := \left\{ T = \varphi_T(\widehat{T}) = \varphi_k(\varphi_{T_\Delta}(\widehat{T})) \mid 1 \le k \le K,\ T_\Delta \in \mathscr{T}_\ell^k \right\}, \quad \forall\, \ell \in \mathbb{N}.$$

Assume that this family of global meshes is shape regular and quasi-uniform. Note that assumption (1), combined with the continuous differentiability of the mappings $(\varphi_k)_{k=1,\dots,K}$ in the compactum Δ^d, implies that it is sufficient to ensure these regularity conditions for each sequence $(\mathscr{T}_\ell^k)_{\ell\in\mathbb{N}}$ separately as far as we keep K fixed.

Finally, let $\mathbb{P} := \mathbb{P}_r(\widehat{T})$ be the space of polynomials of degree r in $\widehat{T}$. Then, for $\ell \in \mathbb{N}$, the parametric finite element space associated with the parametric mesh $\mathscr{T}_\ell$ is

$$\begin{aligned} X_\ell &:= \left\{ v \in \mathscr{C}^0(\Omega) \mid \forall\, T \in \mathscr{T}_\ell\ \exists\, w \in \mathbb{P} :\ v(\boldsymbol{x}) = w(\varphi_T^{-1}(\boldsymbol{x})),\ \forall\, \boldsymbol{x} \in T \right\} \\ &= \left\{ v \in \mathscr{C}^0(\Omega) \mid v \circ \varphi_T \in \mathbb{P},\ \forall\, T \in \mathscr{T}_\ell \right\}. \end{aligned} \tag{3}$$

Note that, in principle, the above definition makes sense for any reasonable set of finite element transformations $(\varphi_T)_{T\in\mathscr{T}_\ell}$. In case the mappings are constructed as in (2) via the high order parameterization from [15], this is a "superparametric" concept if the degree r is small. This is in contrast to the subparametric or isoparametric finite elements which are usually considered in the literature; see [3].

From a practical point of view, virtually every kind of parameterization can be employed with the following qualification. For an efficient assembly of the stiffness matrix and the right hand side via sufficiently accurate (at best exact) numerical quadrature, the derivatives of the resulting finite element transformations (2) and the mappings themselves must be easy to evaluate; see, e.g., [1].

Discretization of Signorini's Problem

Let us now apply the above concept to a contact problem in elasticity to find the deformation of a linear elastic body Ω in contact with a rigid obstacle. For this purpose, let the boundary be decomposed into pairwise disjoint parts: $\partial\Omega = \overline{\Gamma}_D \cup \overline{\Gamma}_N \cup \overline{\Gamma}_C$. Assume that the Dirichlet boundary Γ_D is of positive Lebesgue measure in dimension $d-1$. Moreover, the condition $\overline{\Gamma}_C \cap \overline{\Gamma}_D = \emptyset$ may hold.

Let $\boldsymbol{n}$ be the outer normal vector field on $\partial\Omega \in \mathscr{C}^1$; the initial gap to the rigid obstacle in this direction is given as a function $g : \Gamma_C \to \mathbb{R}_{\ge 0}$. Then, for sufficiently

smooth prescribed volume and surface force densities $\boldsymbol{f} = (f_i)$ and $\boldsymbol{p} = (p_i)$, the displacement field $\boldsymbol{u} : \Omega \to \mathbb{R}^d$ solves the boundary value problem

$$\begin{aligned} -\sigma_{ij}(\boldsymbol{u})_{,j} &= f_i \quad \text{in } \Omega, \\ \boldsymbol{u} &= \boldsymbol{0} \quad \text{on } \Gamma_D, \\ \sigma_{ij}(\boldsymbol{u})n_j &= p_i \quad \text{on } \Gamma_N, \\ \boldsymbol{u}\cdot\boldsymbol{n} &\le g \quad \text{on } \Gamma_C, \end{aligned} \tag{4}$$

where $\sigma_{ij}(\boldsymbol{u}) = A_{ijlm}u_{l,m}$ are the stresses and $\mathbf{A} = (A_{ijlm})$ is Hooke's tensor. The existence of a unique weak solution follows from Lions' and Stampacchia's lemma.

We use the vector-valued parametric finite element space $\boldsymbol{X}_\ell := (X_\ell)^d$ defined by (3) with $r = 1$ and denote the set of nodes by $\mathcal{N}_\ell$. As usual, the non-penetration conditions on the possible contact boundary Γ_C are merely enforced at the potential contact nodes $\mathcal{N}_\ell^C = \mathcal{N}_\ell \cap \Gamma_C$; see below. Then, a discretization of Signorini's problem (4) with one-sided constraints is obtained by specifying a variational inequality

$$\text{find } \boldsymbol{u}_\ell \in \boldsymbol{K}_\ell \text{ such that } a(\boldsymbol{u}_\ell, \boldsymbol{v} - \boldsymbol{u}_\ell) \ge f(\boldsymbol{v} - \boldsymbol{u}_\ell),\ \forall \boldsymbol{v} \in \boldsymbol{K}_\ell, \tag{5}$$

on a suitable set of admissible displacements

$$\boldsymbol{K}_\ell := \left\{ \boldsymbol{v} \in \boldsymbol{X}_\ell \,|\, \boldsymbol{v} = \boldsymbol{0} \text{ on } \Gamma_D,\ (\boldsymbol{v}\cdot\boldsymbol{n})(p) \le g(p),\ \forall p \in \mathcal{N}_\ell^C \right\}.$$

In the discrete variational inequality (5), the (bi-)linear forms a and f representing the elastic energy and the applied forces, respectively, are given by $a(\boldsymbol{u}, \boldsymbol{v}) := \int_\Omega A_{ijlm}u_{l,m}v_{i,j}\,d\boldsymbol{x}$ and $f(\boldsymbol{v}) := \int_\Omega f_i v_i\,d\boldsymbol{x} + \int_{\Gamma_N} p_i v_i\,d\boldsymbol{a}$.

Although, from a modeling point of view, as much geometric information as possible should be used for an accurate description of contact phenomena, we remark that a strong pointwise non-penetration condition everywhere on Γ_C is usually not suitable for the variational formulation on which the (parametric) finite element method relies. Besides, a decoupled set of constraints is preferable for a variety of reasons. The common remedy is to prescribe the contact constraints with respect to a suitable cone of Lagrange multipliers. This requires the introduction of appropriate sets of functionals in $(H^{\frac{1}{2}}(\Gamma_C))'$. To retain inequality constraints which can be enforced merely by looking at the nodes, one can employ discontinuous test spaces described, e.g., in [7].

The quality of a priori error estimates for the above discretization certainly depends on a number of aspects which have to be examined more closely. Beside regularity assumptions for the continuous solution, the balance of the primal degrees of freedom and the constraints by means of an inf-sup condition and certain properties of the parameterization, e.g., the regularity (1), influence the error analysis.

3 Monotone Multigrid Method for Parametric Elements

Similarly to some of the approaches reviewed in [5, Chap. 4], the scale of parametric finite element spaces constitutes an adjusted discretization technique which allows for an almost straightforward application of multilevel ideas. In this section, we

examine the constructed space hierarchy, which we presume to possess the required approximation properties, and the corresponding natural transfer operators in a little more detail.

For the solution of the discrete variational inequality, we propose a monotone multigrid method [12]; see [13] for an overview of this and other solution strategies for contact problems and more references. Here, the non-penetration conditions at the potential contact nodes are treated by a non-linear block Gauß–Seidel smoother at the finest level L. Let $\tilde{\boldsymbol{u}} \in \boldsymbol{K}_L$ be a preliminary approximate solution (i.e., a current admissible iterate). Then, in the next step, a linear multilevel preconditioner depending on $\tilde{\boldsymbol{u}}$ is employed, which acts only on the space $\{\boldsymbol{v} \in \boldsymbol{X}_L \,|\, (\boldsymbol{v} \cdot \boldsymbol{n})(p) = 0,\ \forall\, p \in \mathcal{N}_L^C \text{ with } (\tilde{\boldsymbol{u}} \cdot \boldsymbol{n})(p) = g(p)\}$. The construction of the required coarse spaces from the spaces $(\boldsymbol{X}_\ell)_{\ell<L}$ involves local modifications of the coarse level matrices resulting from recursively truncated basis functions; see, e.g., [13].

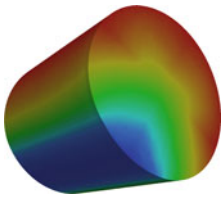
By construction, the spaces defined by (3) are nested. This is an immediate consequence of the fact that the parameterization is fixed and does not change with the index ℓ. Still, let us formulate this statement in the following lemma and give an elementary proof of the assertion.

Lemma 1. *The parametric finite element spaces $(X_\ell)_{\ell\in\mathbb{N}}$ are nested.*

Proof. For $\ell \geq 1$, let $v \in X_{\ell-1}$ be arbitrary. Then, for $T \in \mathscr{T}_{\ell-1}$ there is a unique element $T_\Delta \in \mathscr{T}_{\ell-1}^k$ for some $k \in \{1,\dots,K\}$ such that $\varphi_k(T_\Delta) = T$. Let $(T_\Delta^i)_{i=1,\dots,N}$ be the children of T_Δ in $\mathscr{T}_\ell^k$. In general, $1 \leq N \leq 2^d$; in case of standard uniform refinement of the simplices, it is $N = 2^d$. We have the corresponding set of elements $(T^i)_{i=1,\dots,N}$ in $\mathscr{T}_\ell$ with $T^i = \varphi_k(T_\Delta^i)$ for $i \in \{1,\dots,N\}$. By assumption, $v \circ \varphi_T = v \circ \varphi_k \circ \varphi_{T_\Delta} \in \mathbb{P}$. Therefore, it is $v \circ \varphi_{T^i} = v \circ \varphi_k \circ \varphi_{T_\Delta^i} \in \mathbb{P}$ because $T_\Delta^i \subset T_\Delta$ and the finite element transformations are affine. As each element of $\mathscr{T}_\ell$ appears as the child of an element in $\mathscr{T}_{\ell-1}$ in the above fashion, we obtain $v \in X_\ell$. Consequently, $X_{\ell-1} \subset X_\ell$ for all $\ell \geq 1$. □

Therefore, no advanced transfer concepts need to be studied here as the canonical inclusion $\mathscr{I}_{\ell-1}^\ell : X_{\ell-1} \to X_\ell$ is the most natural operator to be used as prolongation. Note that these operators only depend on the logical structure; as in the standard nested case, the representing matrices contain the entries 0, 0.5 and 1 and may be computed from the neighborhood relations in and between the simplicial meshes $(\mathscr{T}_{\ell-1}^k)_{k=1,\dots,K}$ and $(\mathscr{T}_\ell^k)_{k=1,\dots,K}$. This is because the respective multilevel basis is defined via a lift by proceeding as in (3). As a result, for a fixed finest level L, the computation of the matrices $\mathbf{I}_{\ell-1}^\ell \in \mathbb{R}^{|\mathcal{N}_\ell|\times|\mathcal{N}_{\ell-1}|}$ for $\ell \in \{1,\dots,L\}$ between the nested spaces $(X_\ell)_{\ell=0,\dots,L}$ does not need the parameterization. However, the computation of the outer normals $(\boldsymbol{n}(p))_{p\in\mathcal{N}_L^C}$ and also of the values $(g(p))_{p\in\mathcal{N}_L^C}$ for the prescription of the contact constraints may require access to the mappings $(\varphi_k)_{k=1,\dots,K}$.

We anticipate that the constructed coarse spaces have the desired multilevel approximation properties. More precisely, under mild assumptions on the employed parameterization mappings $(\varphi_k)_{k=1,\dots,K}$, the relevant Jackson- and Bernstein-type inequalities transfer from the standard finite element spaces to the parametric spaces; see also [9].



L	#elements	#dof	#steps	$\tilde{\rho}$	$\lvert\mathscr{A}_L\rvert$
0	96	107	8 (2)	0.032	3
1	768	615	10 (3)	0.031	15
2	6,144	3,915	11 (4)	0.065	58
3	49,152	27,795	13 (6)	0.091	199
4	393,216	209,187	14 (6)	0.102	753
5	3,145,728	1,622,595	15 (8)	0.114	2,984

Fig. 2. Contact problem of a parameterized cylinder with a rigid obstacle shaped like a broad channel. The *colors* indicate the displacement in $\boldsymbol{e}_3$-direction. Problem (5) is solved by a conjugate gradient method preconditioned by the monotone multigrid method ($\mathscr{V}(3,3)$-cycle)

Finally, we point out that no modifications are necessary in the code of the solver provided that the local normal/tangential coordinate systems can be computed from the parameterization. Consequently, a monotone multigrid method can be employed for contact problems discretized with parametric finite elements in the quite straightforward way outlined above. Figure 2 shows a numerical example illustrating the performance of the method for $d = 3$. The number of active nodes where the constraints are binding is denoted by $|\mathscr{A}_L|$. We report on the asymptotic convergence rate $\tilde{\rho}$ of a conjugate gradient method preconditioned by the monotone multigrid method ($\mathscr{V}(3,3)$-cycle). Starting with the initial iterate zero at each refinement level (i.e., no nested iteration), we list the number of total steps needed to reduce the norm of the residual to less than 10^{-10}. The count of included non-linear steps is given in brackets (e.g., for $L = 5$, the active set is found after 8 of the 15 cycles such that the remaining 7 steps are linear). Note that the pcg error reduction rate $\tilde{\rho}$ corresponds to this linear iteration phase where the active set has already been identified.

4 Conclusion

The results described in this paper certainly have preliminary character; the performance of the presented algorithms needs to be studied in more detail. This is work in progress. However, the experiments so far show that (monotone) multigrid methods based on parametric finite elements work as expected; see Fig. 2. Still, the effort of constructing a (high order) parameterization by the methodology developed in [15] especially pays if there is also a considerable gain on the modeling side. Here, the effect of this special resolution of the boundary on the discrete approximation of contact phenomena or general boundary effects needs to be investigated more closely.

Acknowledgments The authors would like to thank Helmut Harbrecht and Maharavo Randrianarivony for bringing this topic to their attention. Moreover, we acknowledge the latter for providing his code for the tetrahedral transfinite interpolation described in [15]. The valuable assistance of Lukas Döring in the implementation of a flexible interface of the parameterization concept to our finite element code is appreciated. This work was supported by the Bonn International Graduate School in Mathematics and the Ford University Research Program.

Bibliography

[1] S. Bartels, C. Carstensen, and A. Hecht. P2Q2Iso2D = 2D isoparametric FEM in Matlab. *J. Comput. Appl. Math.*, 192(2):219–250, 2006.

[2] A. Chernov, M. Maischak, and E.P. Stephan. A priori estimates for *hp* penalty BEM for contact problems in elasticity. *Comput. Methods Appl. Mech. Engrg.*, 196(37–40):3871–3880, 2007.

[3] P.G. Ciarlet. *The Finite Element Method for Elliptic Problems*. North-Holland, Amsterdam, 1978.

[4] W. Dahmen, C.A. Micchelli, and H.P. Seidel. Blossoming begets B-spline bases built better by B-patches. *Math. Comput.*, 59(199):97–115, 1992.

[5] T. Dickopf. *Multilevel Methods Based on Non-Nested Meshes*. PhD thesis, University of Bonn, 2010. http://hss.ulb.uni-bonn.de/2010/2365.

[6] T. Dickopf and R. Krause. Efficient simulation of multi-body contact problems on complex geometries: a flexible decomposition approach using constrained minimization. *Int. J. Numer. Methods Engrg.*, 77(13):1834–1862, 2009.

[7] B. Flemisch and B. Wohlmuth. Stable Lagrange multipliers for quadrilateral meshes of curved interfaces in 3d. *Comput. Methods Appl. Mech. Engrg.*, 196(8):1589–1602, 2007.

[8] W.J. Gordon and C.A. Hall. Transfinite element methods: blending-function interpolation over arbitrary curved element domains. *Numer. Math.*, 21(2):109–129, 1973.

[9] H. Harbrecht. A finite element method for elliptic problems with stochastic input data. *Appl. Numer. Math.*, 60(3):227–244, 2010.

[10] H. Harbrecht and M. Randrianarivony. From computer aided design to wavelet BEM. *Comput. Visual. Sci.*, 13(2):69–82, 2010.

[11] T.J.R. Hughes, J.A. Cottrell, and Y. Bazilevs. Isogeometric analysis: CAD, finite elements, NURBS, exact geometry and mesh refinement. *Comput. Methods Appl. Mech. Engrg.*, 194(39–41):4135–4195, 2005.

[12] R. Kornhuber. *Adaptive Monotone Multigrid Methods for Nonlinear Variational Problems*. Teubner, Stuttgart, 1997.

[13] R. Krause. On the multiscale solution of constrained minimization problems. In U. Langer et al., editor, *Domain Decomposition Methods in Science and Engineering XVII*, volume 60 of *Lect. Notes Comput. Sci. Eng.*, pages 93–104. Springer, 2008.

[14] X.J. Luo, M.S. Shephard, J.F. Remacle, R.M. O'Bara, M.W. Beall, B. Szabó, and R. Actis. p-version mesh generation issues. In *Proceedings of the 11th International Meshing Roundtable*, pages 343–354. 2002.

[15] M. Randrianarivony. Tetrahedral transfinite interpolation with B-patch faces: construction and regularity. INS Preprint No. 0803. University of Bonn, 2008.

[16] P. Seshaiyer and M. Suri. Uniform *hp* convergence results for the mortar finite element method. *Math. Comput.*, 69(230):521–546, 2000.

[17] M. Zlámal. The finite element method in domains with curved boundaries. *Int. J. Numer. Methods Engrg.*, 5(3):367–373, 1973.

TFETI Scalable Solvers for Transient Contact Problems

T. Kozubek, Z. Dostál, T. Brzobohatý, A. Markopoulos, and O. Vlach

Dept. of Appl. Math., VSB-Technical University Ostrava, Czech Republic
tomas.kozubek@vsb.cz, zdenek.dostal@vsb.cz, tomas.brzobohaty@vsb.cz,
alexandros.markopoulos@vsb.cz, oldrich.vlach2@vsb.cz

Summary. We review our results obtained by application of the TFETI domain decomposition method to implement the time step of the Newmark scheme for the solution of transient contact problems without friction. If the ratio of the decomposition and discretization parameters is kept uniformly bounded as well as the ratio of the time and space discretization, then the cost of the time step is proved to be proportional to the number of nodal variables. The algorithm uses our MPRGP algorithm for the solution of strictly convex bound constrained quadratic programming problems with optional preconditioning by the conjugate projector to the subspace defined by the trace of the rigid body motions on the artificial subdomain interfaces. The optimality relies on our results on quadratic programming, the theory of the preconditioning by a conjugate projector for nonlinear problems, and the classical bounds on the spectrum of the mass and stiffness matrices. The results are confirmed by numerical solution of 3D transient contact problems.

1 Introduction

The transient multibody contact problems are important in many applications arising in mechanical or civil engineering. However, it is not easy to provide a useful solution to realistic problems. The reasons include the lack of smoothness, which puts high demand on the construction of effective time discretization schemes, the strong nonlinearity arising from the non-interpenetration boundary conditions, and large dimension of the problems resulting from the space discretization. These complications stimulated extensive research activities both from the theoretical point of view (see, e.g., [4]), or the numerical point of view (see, e.g., [10], or [11]).

Numerical solution of transient contact problems usually comprises several steps. Starting from a week formulation of the conditions of equilibrium and boundary conditions, the problem is first discretized in space by the finite element method in a similar way as the related static problem. The resulting semidiscrete problem is then discretized by a suitable time discretization scheme. The time integration requires a special attention to guarantee stability of the algorithm and to avoid non-physical oscillations that result from application of the standard time discretization methods for unconstrained problems. Such schemes were proposed by many authors (see [6,

R. Bank et al. (eds.), *Domain Decomposition Methods in Science and Engineering XX*, Lecture Notes in Computational Science and Engineering 91,
DOI 10.1007/978-3-642-35275-1_38,

7, 9, 10]). In our approach, we use a combination of the standard finite element space discretization with the contact stabilized Newmark scheme introduced by Krause and Walloth [9] that reduces the solution of the transient contact problem to a sequence of strictly convex quadratic programming (QP) problems with inequality constraints that describe the non-interpenetration conditions.

The final step amounts to the solution of QP problems of large dimension, possibly with millions of nodal variables and many inequality constraints. In this paper we propose to resolve the auxiliary problems by our variant of the FETI domain decomposition method called TFETI (total finite element tearing and interconnecting, Dostál et al. [1]). Our research has been motivated by our recent results in development of optimal algorithms for the frictionless static problems [1] that combine effective FETI preconditioning of both linear and nonlinear steps with our algorithms for the solution of bound constrained QP problems [3]. An important feature of our QP algorithms is the error estimate in terms of the bound on the condition number of the Hessian matrix of the cost function.

2 Transient Contact Problem and Its Discretization Using TFETI

The starting point of our exposition is the discretized transient multibody contact problem resulting from application of our TFETI domain decomposition. The reason is that a little is known about the solvability of the weak formulation of the transient contact problem (see, e.g., [4]), so we shall assume in what follows that its solution $\mathbf{u}$ exists. Moreover, we shall assume that $\mathbf{u}$ is sufficiently smooth so that $\ddot{\mathbf{u}}$ exists in some reasonable sense and can be approximated by finite differences. More specific choice of the solution space can be found, e.g., in [4] or in [6].

To discretize the multibody contact problem using TFETI, we tear each body from the part of the boundary with the Dirichlet boundary conditions, decompose each body into subdomains, assign each subdomain a unique number, and introduce new "gluing" conditions on the artificial subdomain interfaces and on the boundaries with imposed Dirichlet conditions. We denote the subdomains and their number by Ω^p and s, respectively. The gluing conditions require continuity of the displacements and of their normal derivatives across the subdomain interfaces. The procedure is the same as that for the static problem, [1].

Using finite element discretization in space we get the following semidiscrete problem at time τ

$$\mathbf{M}\ddot{\mathbf{u}} + \mathbf{K}\mathbf{u} = \mathbf{f} - \mathbf{B}_I^T \boldsymbol{\lambda}_I^T - \mathbf{B}_E^T \boldsymbol{\lambda}_E, \tag{1}$$

$$\mathbf{B}_I\mathbf{u} \le \mathbf{c}_I, \quad \mathbf{B}_E\mathbf{u} = \mathbf{c}_E, \quad \boldsymbol{\lambda}_I \ge \mathbf{o}, \quad \boldsymbol{\lambda}^T(\mathbf{B}\mathbf{u} - \mathbf{c}) = 0, \tag{2}$$

with the discrete Newton equation of motion (1) and the equality and inequality constraints (2) resulting from the gluing, Dirichlet, and non-interpenetration conditions enforced by Lagrange multipliers.

The TFETI based finite element semi-discretization in space of the subdomains Ω^p, $p = 1, \ldots, s$, results in the block diagonal stiffness matrix $\mathbf{K} = \mathrm{diag}\,(\mathbf{K}_1, \ldots, \mathbf{K}_s)$

of the order n with the sparse positive semidefinite diagonal blocks $\mathbf{K}_p$ that correspond to the subdomains Ω^p. The same structure has a positive definite mass matrix $\mathbf{M} = \mathrm{diag}\,(\mathbf{M}_1, \ldots, \mathbf{M}_s)$. The decomposition induces also the block structure of the vector of nodal forces $\mathbf{f} = \mathbf{f}_\tau \in \mathbb{R}^n$ at time τ and the vector of nodal displacements $\mathbf{u} = \mathbf{u}_\tau \in \mathbb{R}^n$ at time τ.

The matrix $\mathbf{B}_I \in \mathbb{R}^{m_I \times n}$ and the vector $\mathbf{c}_I \in \mathbb{R}^{m_I}$ describe the linearized non-interpenetration conditions and the matrix $\mathbf{B}_E \in \mathbb{R}^{m_E \times n}$ and the vector $\mathbf{c}_E \in \mathbb{R}^{m_E}$ enforce the prescribed zero displacements on the part of the boundary with imposed Dirichlet condition and the continuity of the displacements across the auxiliary interfaces.

Finally, $\boldsymbol{\lambda}_I \in \mathbb{R}^{m_I}$ and $\boldsymbol{\lambda}_E \in \mathbb{R}^{m_E}$ denote the components of the vector of Lagrange multipliers $\boldsymbol{\lambda} = \boldsymbol{\lambda}_\tau \in \mathbb{R}^m$, $m = m_I + m_E$ at time τ. We use the notation

$$\boldsymbol{\lambda} = \begin{bmatrix} \boldsymbol{\lambda}_I \\ \boldsymbol{\lambda}_E \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} \mathbf{B}_I \\ \mathbf{B}_E \end{bmatrix}, \quad \text{and} \quad \mathbf{c} = \begin{bmatrix} \mathbf{c}_I \\ \mathbf{c}_E \end{bmatrix}. \tag{3}$$

For the time discretization, we use the contact-stabilized Newmark scheme introduced by Krause and Walloth [9] with the regular partition of the time interval $[0, T]$, $0 = \tau_0 < \tau_1 \ldots < \tau_{n_T} = T$, $\tau_k = k\Delta$, $\Delta = T/n_T$, $k = 0, \ldots, n_T$. The scheme assumes that the acceleration vector is split at time τ_k into two components

$$\ddot{\mathbf{u}}_k = \ddot{\mathbf{u}}_k^{int} + \ddot{\mathbf{u}}_k^{con}, \quad \ddot{\mathbf{u}}_k^{int} = \mathbf{M}^{-1}\left(\mathbf{f}_k - \mathbf{K}\mathbf{u}_k\right), \text{ and } \ddot{\mathbf{u}}_k^{con} = -\mathbf{M}^{-1}\mathbf{B}^T\boldsymbol{\lambda}_k. \tag{4}$$

We obtain the solution algorithm in the form

Algorithm 2.1 Contact-stabilized Newmark algorithm.

Step 0. {*Initialization*}
Set $\mathbf{u}_0$, $\dot{\mathbf{u}}_0$, $\widetilde{\mathbf{K}} = \frac{4}{\Delta^2}\mathbf{M} + \mathbf{K}$, $T > 0$, $n_T \in \mathbb{N}$, and $\Delta = T/n_T$.

for $k = 0, \ldots, n_T - 1$ **do**

Step 1. {*Predictor displacement computation*}
$$\min \left[\tfrac{1}{2} \left(\mathbf{u}_{k+1}^{pred}\right)^T \mathbf{M}\mathbf{u}_{k+1}^{pred} - \left(\mathbf{M}\mathbf{u}_k + \Delta\mathbf{M}\dot{\mathbf{u}}_k - \mathbf{B}^T\boldsymbol{\lambda}_k^{pred}\right)^T \mathbf{u}_{k+1}^{pred} \right]$$
subject to $\mathbf{B}_I\mathbf{u}_{k+1}^{pred} \le \mathbf{c}_I$, and $\mathbf{B}_E\mathbf{u}_{k+1}^{pred} = \mathbf{c}_E$

Step 2. {*Contact-stabilized displacement computation*}
$$\min \left[\tfrac{1}{2}\mathbf{u}_{k+1}^T\widetilde{\mathbf{K}}\mathbf{u}_{k+1} - \left(\tfrac{4}{\Delta^2}\mathbf{M}\mathbf{u}_{k+1}^{pred} - \mathbf{K}\mathbf{u}_k + \mathbf{f}_k + \mathbf{f}_{k+1} - \mathbf{B}^T\boldsymbol{\lambda}_k\right)^T \mathbf{u}_{k+1} \right]$$
subject to $\mathbf{B}_I\mathbf{u}_{k+1} \le \mathbf{c}_I$ and $\mathbf{B}_E\mathbf{u}_{k+1} = \mathbf{c}_E$

Step 3. {*Velocity evaluation*}
$$\dot{\mathbf{u}}_{k+1} = \dot{\mathbf{u}}_k + \tfrac{2}{\Delta}\left(\mathbf{u}_{k+1} - \mathbf{u}_{k+1}^{pred}\right)$$

end

The matrix $\widetilde{\mathbf{K}}$ introduced in *Step 0* is called an *effective stiffness matrix*. Let us note that we omit the factor '1/2' in the term $\mathbf{B}^T\boldsymbol{\lambda}_k^{pred}$ in the predictor step.

3 Optimal Solver with Bound on the Condition Number of the Hessian of the Dual Energy Function

The favorable distribution of the spectrum of the mass matrix $\mathbf{M}$ is sufficient to implement Step 1 by using the dual theory and the standard MPRGP algorithm described in [3] with asymptotically linear complexity. To develop an optimal algorithm for Step 2, we shall distinguish two cases. If the time steps are sufficiently short, then the effective stiffness matrix can be considered as a perturbation of the well conditioned mass matrix, so it is enough to use again our MPRGP algorithm to prove the numerical scalability and demonstrate it by numerical experiments. On the other hand, if we use longer time steps, the effective stiffness matrix has very small eigenvalues which obviously correspond to the eigenvectors that are near the kernel of $\mathbf{K}$. This observation was fully exploited for linear problems by Farhat et al. [5] who used the conjugate projectors to the natural coarse grid to achieve scalability with respect to the time step. Unfortunately, this idea can not be applied in full extent to the contact problems as we do not know a priori which boundary conditions are applied to the subdomains associated with the contact interface. However, we can still define the preconditioning by the trace of the rigid body motions on the artificial subdomain interfaces. To implement this observation, we use our preconditioning by conjugate projector for partially constrained strictly convex quadratic programming problems of the form

$$\min_{\boldsymbol{\lambda}} \ \frac{1}{2}\boldsymbol{\lambda}^T\widetilde{\mathbf{F}}\boldsymbol{\lambda} \ - \ \boldsymbol{\lambda}^T\mathbf{d} \text{ subject to } \boldsymbol{\lambda}_{\mathscr{I}} \geq \mathbf{o} \tag{5}$$

which arises directly from the application of the dual theory on the problem in Step 2 of Algorithm 2.1. Such a method complies with our MPRGP-P algorithm for the solution of strictly convex bound constrained problems described in [3]. We keep the iterations in the subspace with the solution which is defined by the trace of the rigid body motions on the artificial interfaces between subdomains excluding the contact interface. Even though the necessity to keep the coarse grid away from the contact interface prevented us from proving the optimality with respect to the time step, we give the proof of optimality of our algorithm provided the ratio of the time step and the space discretization parameter is kept uniformly bounded and show that the optimality can be observed by numerical experiments (see [2] for details). Moreover, MPRGP-P algorithm has the rate of convergence in terms of the norm of the projected gradient and the bound on the condition number of the Hessian matrix of the cost functional. Therefore all we need to guarantee optimality is a uniform bound on the condition number of the Hessian.

In [2], we used the standard arguments to prove the following lemma which gives the required bound.

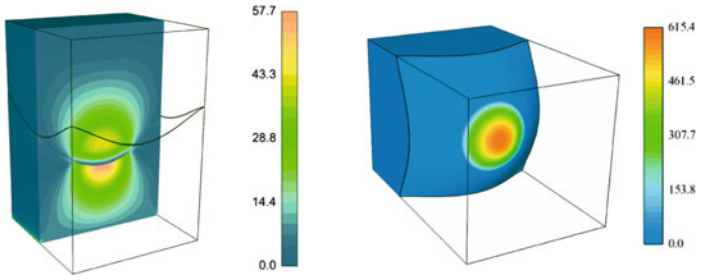
Lemma 1. *Let* $B_1\|\boldsymbol{\lambda}\|^2 \leq \|\mathbf{B}^T\boldsymbol{\lambda}\|^2 \leq B_2\|\boldsymbol{\lambda}\|^2$ *and let the elements have a regular shape and size. Then*

$$C_1\frac{h^2\Delta^2}{h^d\,(h^2+\Delta^2)}\|\boldsymbol{\lambda}\|^2 \leq \boldsymbol{\lambda}^T\widetilde{\mathbf{F}}\boldsymbol{\lambda} \leq C_2\frac{\Delta^2}{h^d}\|\boldsymbol{\lambda}\|^2, \tag{6}$$

with constants B_1, B_2, C_1, and C_2 independent of h, H, and Δ. Moreover, if $C > 0$ is any constant, then for any $0 < \Delta \leq Ch$ the condition number $\kappa(\widetilde{\mathbf{F}})$ satisfies $\kappa(\widetilde{\mathbf{F}}) \leq \frac{C_2}{C_1}(1+C^2)$.

4 Numerical Experiments

The described algorithms were implemented in `MatSol` library [8] developed in Matlab environment and tested on the solution of 3D frictionless transient contact problems. For all computations we used the HP Blade system, model BLc7000 and as parallel programming environment we used Matlab Distributed Computing Engine. All the computations were carried out with the relative stopping tolerance $\varepsilon = 10^{-4}$.



a) Von Mises stress in τ_1 [MPa] b)Contact pressure in τ_1 [MPa]

Fig. 1. Results of 3D benchmark

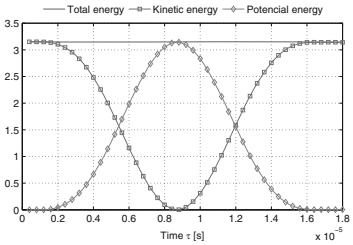


Fig. 2. Energy conservation (ton·mm^2·s^{-2})

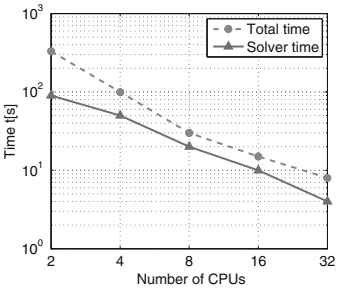


Fig. 3. Parallel scalability

3D impact problem

Our first academic benchmark is a 3D impact between the curved 3D elastic boxes of size 10 (mm) depicted in Fig. 1. Material constants are defined by the Young modulus $E = 2.1 \cdot 10^5$ (MPa), the Poisson ratio $\nu = 0.3$, and the density $\rho = 7.85 \cdot 10^{-9}$ (ton/mm^3). The initial gap between the curved boxes is set to 0.001 (mm). We prescribe the initial velocity $-1{,}000$ (mm/s) on the upper body in the x_3 direction. The

upper body is floating in space and the lower body is fixed along the bottom side. The linearized non-interpenetration condition was imposed on the contact interface. For the time discretization, we use Algorithm 2.1 with the constant time step $\Delta = 4 \cdot 10^{-7}$ and solve the impact of bodies in the time interval $\tau = [0, 45\Delta]$.

The von Mises stress distribution and the normal contact pressure along the contact interface in time $\tau_1 = 22\Delta$ are depicted in Figs. 1a, b, respectively. The energy development is shown in Fig. 2. We can see the constant total energy curve as expected.

In Table 1, we report the numerical scalability of our algorithm for the constant time step $\Delta_1 = 1 \cdot 10^{-3}$ and $\Delta_2 = 1 \cdot 10^{-5}$ and with or without conjugate projectors. We kept $H/h = 10$. Moreover, in last two lines of the table, we report the same characteristics but with the time step dependent on the discretization step h, i.e., $\Delta_{1,h} = 3h\Delta_1$.

We can observe that the number of matrix-vector multiplications, the most expensive component of our algorithm, stays constant for the smaller time step Δ_2 as expected and increases only mildly in agreement with the theory for the case of the larger time step Δ_1 if we use conjugate projectors. If we simultaneously choose the time step Δ proportional to h, i.e., $\Delta = \Delta_h$, then the number of matrix-vector multiplications stays the same as predicted by the theory.

Parallel scalability of our algorithm is depicted in Fig. 3, where we keep the number of elements fixed and increase the number of CPUs (subdomains).

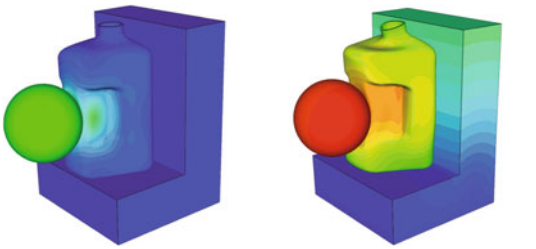
Number of subdomains		16	54	128	250
Primal variables		196 608	663 552	1 572 864	3 072 000
Dual variables		21 706	81 652	214 699	443 920
		Hessian multiplications			
MPRGP	Δ_1	67	86	113	191
MPRGP - P	Δ_1	60	67	85	112
MPRGP	Δ_2	39	40	40	42
MPRGP - P	Δ_2	40	40	40	42
MPRGP	$\Delta_{1,h}$	67	72	76	78
MPRGP - P	$\Delta_{1,h}$	60	63	67	69

Table 1. Numerical scalability of 3D impact problem - Δ constant or dependent on h

Impact of three bodies

We have also tested our algorithms on the impact of three bodies. We considered the transient analysis of three elastic bodies in mutual contact (see Fig. 4). We prescribe the initial velocity 5,000 (mm/s) on the sphere in the x_1 direction. The L-shape body is fixed along the bottom side. Material constants are defined by the Young modulus $E = 2.1 \cdot 10^3$ (MPa), the Poisson ratio $\nu = 0.3$, and the density $\rho = 6 \cdot 10^{-9}$ (ton/mm^3). For the time discretization, we use the constant time step $\Delta = 1 \cdot 10^{-3}$ (s) and solve the impact of bodies in the time interval $\tau = [0, 150\Delta]$ (s). The total displacement in times $\tau_1 = 20\Delta$ and $\tau_2 = 80\Delta$ (s) of the problem discretized by $1.2 \cdot 10^5$

primal and $8.5 \cdot 10^3$ dual variables and decomposed into 32 subdomains using METIS is depicted in Fig. 4.



a) Total displacement in τ_1 b) Total displacement in τ_2

Fig. 4. Impact of bodies in time

Acknowledgments The work is supported by the project of Ministry of Education of the Czech Republic MSM6198910027 and by the project 101/08/0574 of the Grant Agency of the Czech Republic.

Bibliography

[1] Z. Dostál, T. Kozubek, V. Vondrák, T. Brzobohatý, and A. Markopoulos. Scalable TFETI algorithm for the solution of multibody contact problems of elasticity. *Internat. J. Numer. Methods Engrg.*, 82(11):1384–1405, 2010. ISSN 0029-5981.

[2] Z. Dostál, T. Kozubek, T. Brzobohatý, A. Markopoulos, and O. Vlach. Scalable TFETI with preconditioning by conjugate projector for transient frictionless contact problems of elasticity. submitted., 2011.

[3] Zdeněk Dostál. *Optimal quadratic programming algorithms*, volume 23 of *Springer Optimization and Its Applications*. Springer, New York, 2009. ISBN 978-0-387-84805-1. With applications to variational inequalities.

[4] Christof Eck, Jiří Jarušek, and Miroslav Krbec. *Unilateral contact problems*, volume 270 of *Pure and Applied Mathematics (Boca Raton)*. Chapman & Hall/CRC, Boca Raton, FL, 2005. ISBN 978-1-57444-629-6; 1-57444-629-0. doi: 10.1201/9781420027365. URL http://dx.doi.org/10.1201/9781420027365. Variational methods and existence theorems.

[5] C. Farhat, P.S. Chen, and J. Mandel. A scalable lagrange multiplier based domain decomposition method for time-dependent problems. *Internat. J. Numer. Methods Engrg.*, 38(22):3831–3853, 1995. ISSN 0029-5981.

[6] C. Hager and B. I. Wohlmuth. Analysis of a space-time discretization for dynamic elasticity problems based on mass-free surface elements. *SIAM J. Numer.*

Anal., 47(3):1863–1885, 2009. ISSN 0036-1429. doi: 10.1137/080715627. URL http://dx.doi.org/10.1137/080715627.

[7] Houari Boumediène Khenous, Patrick Laborde, and Yves Renard. Mass redistribution method for finite element contact problems in elastodynamics. *Eur. J. Mech. A Solids*, 27(5):918–932, 2008. ISSN 0997-7538. doi: 10.1016/j.euromechsol.2008.01.001. URL http://dx.doi.org/10.1016/j.euromechsol.2008.01.001.

[8] T. Kozubek, A. Markopoulos, T. Brzobohatý, R. Kučera, V. Vondrák, and Z. Dostál. Matsol - matlab efficient solvers for problems in engineering. "http://matsol.vsb.cz/", 2009.

[9] Rolf Krause and Mirjam Walloth. A time discretization scheme based on Rothe's method for dynamical contact problems with friction. *Comput. Methods Appl. Mech. Engrg.*, 199(1-4):1–19, 2009. ISSN 0045-7825. doi: 10.1016/j.cma.2009.08.022. URL http://dx.doi.org/10.1016/j.cma.2009.08.022.

[10] Tod A. Laursen. *Computational contact and impact mechanics*. Springer-Verlag, Berlin, 2002. ISBN 3-540-42906-9. Fundamentals of modeling interfacial phenomena in nonlinear finite element analysis.

[11] Peter Wriggers. *Computational contact mechanics*. John Wiley & Sons, Ltd., Chichester, West Sussex, England, 2002.

Model of Imperfect Interfaces in Composite Materials and Its Numerical Solution by FETI Method

Jaroslav Kruis, Jan Zeman, and Pavel Gruber

Department of Mechanics, Faculty of Civil Engineering, Czech Technical University in Prague jk@cml.fsv.cvut.cz, zeman@cml.fsv.cvut.cz, gruber@fsv.cvut.cz

Summary. Analysis of material interfaces in composite materials is in the center of attention of many material engineers. The material interface influences significantly the overall behaviour of composite materials. While the perfect bond on material interface is modelled without larger difficulties, the imperfect bond between different components of composite materials still causes some obstacles. This contribution concentrates on application of the FETI method to description of the imperfect bond.

1 Introduction

The overall behavior of the engineering materials and structures is significantly affected or even dominated by the presence of interfaces, i.e. internal boundaries arising from material discontinuities. Therefore, considerable research efforts within the engineering community have been focused to adequately describe and simulate the interfacial behavior under general loading conditions. A successful approach to this problem is offered by the cohesive zone concept published in reference [3], in which the bulk material is assumed to be damage-free, whereas the interface response is described by means of inelastic damage law. The interface model itself is formulated in terms of displacement jumps and cohesive tractions bridging the interface, with the elastic stiffness as the basic constitutive parameter. Initially, the stiffness is set to a large value (modeling almost perfect bonding) that gradually decreases with increasing load. For the standard displacement-based finite element approximations, this gives a rise to numerical difficulties manifested in oscillations of interfacial tractions for stiff interfaces and non-physical penetration of adjacent bodies for imperfect bonding. The purpose of this contribution is to demonstrate that these limitations can be overcome by duality solvers based on FETI method.

2 Interface Model

The constitutive description adopted in this work is based on the Ortiz-Pandolfi model proposed in [7]. Detailed description of the model of the imperfect material

R. Bank et al. (eds.), *Domain Decomposition Methods in Science and Engineering XX*, Lecture Notes in Computational Science and Engineering 91,
DOI 10.1007/978-3-642-35275-1_39,

interface can be found in reference [2]. The model is based on three state variables, namely the domain displacement field, $\mathbf{u}^{(j)}(\mathbf{x})$, the interfacial displacement jump, $[\![\mathbf{u}^{(i,j)}]\!](\mathbf{x})$, and the interfacial damage parameter, $\omega^{(i,j)}(\mathbf{x})$. The superscript (j) denotes the subdomain number while the two superscripts (i,j) denote the interface between the i-th and j-th subdomains.

The kinematics of the interface is quantified by the normal and tangential component of the displacement jump, provided by

$$[\![u_n^{(i,j)}]\!](\mathbf{x}) = [\![\mathbf{u}^{(i,j)}]\!](\mathbf{x}) \cdot \mathbf{n}^{(j)}(\mathbf{x}), \tag{1}$$

where $\mathbf{n}^{(j)}(\mathbf{x})$ denotes the normal vector and the tangential component is in the form

$$[\![\mathbf{u}_t^{(i,j)}]\!](\mathbf{x}) = [\![\mathbf{u}^{(i,j)}]\!](\mathbf{x}) - [\![u_n^{(i,j)}]\!](\mathbf{x})\mathbf{n}^{(j)}(\mathbf{x}). \tag{2}$$

Note that the non-penetration condition hold, i.e. the normal component must remain non-negative. Following [3], these quantities are combined into an effective opening

$$\delta(\mathbf{x}, [\![\mathbf{u}^{(i,j)}]\!](\mathbf{x})) = \sqrt{[\![u_n^{(i,j)}]\!]^2(\mathbf{x}) + \beta^2 \|[\![\mathbf{u}_t^{(i,j)}]\!](\mathbf{x})\|^2} \tag{3}$$

in which β denotes a constitutive parameter, also called the mode mixity parameter, to be determined. This gives rise to an equivalent effective traction, σ, see [7]. In addition, the state of an interface is quantified by an internal damage variable, ω, with $\omega(\mathbf{x}) = 0$ corresponding to a perfect bonding at $\mathbf{x}$, whereas $\omega(\mathbf{x}) = 1$ indicates a fully damaged interface point.

In order to assemble the functional of energy, several energy densities are needed. The density of internal energy has the form

$$e_{vol}^{(j)}(\mathbf{x}, \mathbf{u}^{(j)}(\mathbf{x})) = \frac{1}{2}\left(\varepsilon(\mathbf{u}^{(j)}(\mathbf{x}))\right)^T \mathbf{D}\varepsilon(\mathbf{u}^{(j)}(\mathbf{x})), \tag{4}$$

where $\varepsilon^{(j)}(\mathbf{u}^{(j)}(\mathbf{x}))$ denotes the strain, $\mathbf{D}$ denotes the stiffness matrix of the material. The internal energy functional can be written as

$$E_{vol}^{(j)}(\mathbf{u}^{(j)}(\mathbf{x})) = \int_{\Omega^{(j)}} e_{vol}^{(j)}(\mathbf{x}, \mathbf{u}^{(j)}(\mathbf{x}))\mathrm{d}\Omega. \tag{5}$$

The potential energy of external forces has the form

$$E_{ext}^{(j)}(\mathbf{u}^{(j)}(\mathbf{x}), t) = -\int_{\Omega^{(j)}} \mathbf{u}^{(j)}(\mathbf{x}) \cdot \mathbf{b}(\mathbf{x}, t)\mathrm{d}\Omega - \int_{\Gamma_t^{(j)}} \mathbf{u}^{(j)}(\mathbf{x}) \cdot \mathbf{t}(\mathbf{x}, t)\mathrm{d}\Gamma, \tag{6}$$

where $\mathbf{b}(\mathbf{x}, t)$ denotes the vector of volume forces, $\mathbf{t}(\mathbf{x}, t)$ denotes the vector of surface traction and $\Gamma_t^{(j)}$ is the part of the boundary of the j-th subdomain where the surface tractions are prescribed. The energy-based description involves the stored energy function defined as

$$e_{int}(\mathbf{x}, [\![\mathbf{u}]\!](\mathbf{x}), \omega(\mathbf{x})) = \frac{1}{2}\frac{G}{\Delta^2}\frac{1-\omega(\mathbf{x})}{\omega(\mathbf{x})}\delta^2, \tag{7}$$

where Δ is the critical interface opening and G is the fracture toughness of an interface. This form is consistent with the linear softening law drawn in Fig. 1. Note that the stiffness associated with a partially damaged interface with the damage parameter, ω, is obtained as a slope of the line $0A$. The energy dissipated by changing the internal variable from ω_1 to ω_2 is given by

$$d = \begin{cases} G(\mathbf{x})(\omega_2(\mathbf{x}) - \omega_1(\mathbf{x})) & \forall \mathbf{x} \in \Gamma_{int} : \omega_1(\mathbf{x}) \leq \omega_2(\mathbf{x}), \\ \infty & otherwise, \end{cases} \tag{8}$$

where the term ∞ refers to the fact that the damage variable cannot decrease during the loading process. The interfacial dissipation distance is defined

$$D(\omega_1(\mathbf{x}), \omega_2(\mathbf{x})) = \int_{\Gamma_{int}} d(\mathbf{x}, \omega_1(\mathbf{x}), \omega_2(\mathbf{x}))\mathrm{d}\Gamma. \tag{9}$$

The interfacial energy functional has the form

$$E_{int}([\![\mathbf{u}]\!](\mathbf{x}), \omega(\mathbf{x})) = \int_{\Gamma_{int}} e_{int}(\mathbf{x}, [\![\mathbf{u}]\!](\mathbf{x}), \omega(\mathbf{x}))\mathrm{d}\Gamma, \tag{10}$$

where Γ_{int} denotes the interface between subdomains.

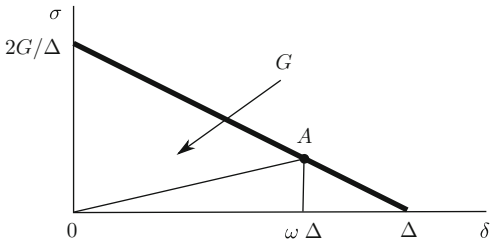


Fig. 1. Interfacial constitutive law

The description of the material interface is based on incremental solution where the state variables at the k-th step $\mathbf{u}_{k-1}(\mathbf{x})$, $[\![\mathbf{u}]\!]_{k-1}(\mathbf{x})$, $\omega_{k-1}(\mathbf{x})$ are known. Then, the energy functional has the form

$$\Pi_k(\mathbf{u}(\mathbf{x}), [\![\mathbf{u}]\!](\mathbf{x}), \omega(\mathbf{x})) = \sum_{j=1}^{n} E_{vol}^{(j)}(\mathbf{u}^{(j)}(\mathbf{x})) + \tag{11}$$

$$\sum_{j=1}^{n} E_{ext}^{(j)}(\mathbf{u}^{(j)}(\mathbf{x})) + E_{int}([\![\mathbf{u}]\!](\mathbf{x}), \omega(\mathbf{x})) + D(\omega_{k-1}(\mathbf{x}), \omega(\mathbf{x}))$$

and the following minimization problem is solved

$$(\mathbf{u}_k(\mathbf{x}), [\![\mathbf{u}]\!]_k(\mathbf{x}), \omega_k(\mathbf{x})) = \arg \min_{(\mathbf{u}(\mathbf{x}), [\![\mathbf{u}]\!](\mathbf{x}), \omega(\mathbf{x}))} \Pi_k(\mathbf{u}(\mathbf{x}), [\![\mathbf{u}]\!](\mathbf{x}), \omega(\mathbf{x})). \tag{12}$$

The discretization of displacements and strains has the form

$$\mathbf{u}^{(j)}(\mathbf{x}) \approx \mathbf{u}_h^{(j)}(\mathbf{x}) = \mathbf{N}_{u,h}^{(j)}(\mathbf{x})\mathbf{u}_h^{(j)}, \tag{13}$$

$$\boldsymbol{\varepsilon}^{(j)}(\mathbf{x}) \approx \boldsymbol{\varepsilon}_h^{(j)}(\mathbf{x}) = \mathscr{B}_{u,h}^{(j)}(\mathbf{x})\mathbf{u}_h^{(j)}, \tag{14}$$

where $\mathbf{N}_{u,h}^{(j)}(\mathbf{x})$ denotes the matrix of basis functions and $\mathscr{B}_{u,h}^{(j)}(\mathbf{x})$ denotes the strain-displacement matrix. The displacement jump is discretized in the form

$$[\![\mathbf{u}^{(i,j)}]\!](\mathbf{x}) \approx [\![\mathbf{u}_h^{(i,j)}]\!](\mathbf{x}) = \mathbf{N}_{[\![u]\!],h}^{(i,j)}(\mathbf{x})[\![\mathbf{u}^{(i,j)}]\!]_h \tag{15}$$

and the damage parameter can be expressed

$$\omega^{(i,j)}(\mathbf{x}) \approx \omega_h^{(i,j)}(\mathbf{x}) = \mathbf{N}_{\omega,h}^{(i,j)}(\mathbf{x})\omega_h^{(i,j)}. \tag{16}$$

After discretization, the functional of energy (11) has the form

$$\begin{aligned}\Pi_k(\mathbf{u}_h, [\![\mathbf{u}]\!]_h, \omega_h) = \frac{1}{2}\sum_{j=1}^{n} {\mathbf{u}_h^{(j)}}^T \mathbf{K}^{(j)}\mathbf{u}_h^{(j)} - \sum_{j=1}^{n} {\mathbf{u}_h^{(j)}}^T \mathbf{f}_h^{(j)} + \\ + \frac{1}{2}[\![\mathbf{u}]\!]_h^T \mathbf{K}_{int}(\omega_h)[\![\mathbf{u}]\!]_h + \omega_h^T \mathbf{p}_h,\end{aligned} \tag{17}$$

where the stiffness matrix has the classical form

$$\mathbf{K}^{(j)} = \int_{\Omega^{(j)}} {\mathscr{B}_{u,h}^{(j)}}^T \mathbf{D}\mathscr{B}_{u,h}^{(j)} \mathrm{d}\Omega \tag{18}$$

and the vector of prescribed forces is defined as

$$\mathbf{f}_h^{(j)} = \int_{\Omega^{(j)}} {\mathbf{N}_{u,h}^{(j)}}^T(\mathbf{x})\mathbf{b}(\mathbf{x})\mathrm{d}\Omega + \int_{\Gamma_t^{(j)}} {\mathbf{N}_{u,h}^{(j)}}^T(\mathbf{x})\mathbf{t}(\mathbf{x},t)\mathrm{d}\Gamma. \tag{19}$$

The stiffness matrix of the interface has the form

$$\mathbf{K}_{int}(\omega_h) = \int_{\Gamma_{int}} \frac{G}{\Delta^2}\left(\frac{1}{\mathbf{N}_{\omega,h}(\mathbf{x})\omega_h} - 1\right)\mathbf{N}_{[\![u]\!],h}^T(\mathbf{x})\beta\mathbf{N}_{[\![u]\!],h}(\mathbf{x})\mathrm{d}\Gamma \tag{20}$$

and the vector $\mathbf{p}_h$ is expressed as

$$\mathbf{p}_h = \int_{\Gamma_{int}} G(\mathbf{x})\mathbf{N}_{\omega,h}(\mathbf{x})\mathrm{d}\Gamma. \tag{21}$$

The minimization (12) is done by the alternate minimization approach which can be written as

$$(\mathbf{u}_k(\mathbf{x}), [\![\mathbf{u}]\!]_k(\mathbf{x}), \omega_k(\mathbf{x})) = \arg\min_{\omega(\mathbf{x})}\left(\min_{(\mathbf{u}(\mathbf{x}),[\![\mathbf{u}]\!](\mathbf{x}))} \Pi_k(\mathbf{u}(\mathbf{x}), [\![\mathbf{u}]\!](\mathbf{x}), \omega(\mathbf{x}))\right). \tag{22}$$

The minimization with respect to $\mathbf{u}(\mathbf{x})$ and $[\![\mathbf{u}(\mathbf{x})]\!]$ is associated with the Lagrangian function in the form

$$L_{k,h}(\mathbf{u}_h, [\![\mathbf{u}]\!]_h, \lambda_h) = \frac{1}{2}\sum_{j=1}^{n} \mathbf{u}_h^{(j)^T} \mathbf{K}^{(j)} \mathbf{u}_h^{(j)} - \sum_{j=1}^{n} \mathbf{u}_h^{(j)^T} \mathbf{f}_h^{(j)} + \tag{23}$$
$$+ \frac{1}{2}[\![\mathbf{u}]\!]_h^T \mathbf{K}_{int}(\omega_h)[\![\mathbf{u}]\!]_h + \lambda_h^T(\mathbf{B}_h\mathbf{u}_h - [\![\mathbf{u}]\!]_h).$$

Note that the displacement jumps $[\![\mathbf{u}]\!]_h$ are subject to the non-penetration condition $\mathbf{B}_h[\![\mathbf{u}]\!]_h \geq 0$. In the current implementation, these constraints are converted to equalities by adopting a simple active set strategy based on the values of the Lagrange multipliers λ_h. There are three stationary conditions

$$\frac{\partial L_{k,h}}{\partial \mathbf{u}_h^{(j)}} = \mathbf{K}^{(j)}\mathbf{u}_h^{(j)} - \mathbf{f}_h^{(j)} + \mathbf{B}_{u,h}^{(j)^T}\lambda_h = \mathbf{0}, \tag{24}$$

$$\frac{\partial L_{k,h}}{\partial \lambda_h} = \sum_{j=1}^{n} \mathbf{B}_{u,h}^{(j)}\mathbf{u}_h^{(j)} - [\![\mathbf{u}]\!]_h = \mathbf{0}, \tag{25}$$

$$\frac{\partial L_{k,h}}{\partial [\![\mathbf{u}]\!]_h} = \mathbf{K}_{int}(\omega_h)[\![\mathbf{u}]\!]_h - \lambda_h = \mathbf{0}. \tag{26}$$

Equation (24) is the equilibrium equation for the j-th subdomain, (25) expresses the interface conditions and (26) defines the relationship between the Lagrange multipliers and the displacement jumps on the interface.

3 FETI Method

This section summarizes the notation and the basic relationships of the FETI method which is a non-overlapping domain decomposition method. More details can be found in references [1, 4] or [5]. The vector of unknowns is denoted by $\mathbf{u}$, the vector of prescribed forces is denoted by $\mathbf{f}$ and the stiffness matrix is denoted by $\mathbf{K}$. Interface conditions for perfect and imperfect interaction have the form

$$\mathbf{B}\mathbf{u} = \begin{pmatrix} \mathbf{B}_c \\ \mathbf{B}_s \end{pmatrix} \mathbf{u} = \begin{pmatrix} \mathbf{0} \\ \mathbf{s} \end{pmatrix} = \mathbf{c}, \tag{27}$$

where $\mathbf{s}$ denotes the jump between subdomain displacements.

After space discretization, the functional of energy has the form

$$\Pi = \Pi(\mathbf{u}, \lambda) = \frac{1}{2}\mathbf{u}^T\mathbf{K}\mathbf{u} - \mathbf{u}^T\mathbf{f} + \lambda^T(\mathbf{B}\mathbf{u} - \mathbf{c}), \tag{28}$$

where λ denotes the vector of Lagrange multipliers.

The interface condition and the solvability condition define the coarse problem

$$\begin{pmatrix} \mathbf{F} & \mathbf{G} \\ \mathbf{G}^T & \mathbf{0} \end{pmatrix} \begin{pmatrix} \lambda \\ \alpha \end{pmatrix} = \begin{pmatrix} \mathbf{d} - \mathbf{c} \\ \mathbf{e} \end{pmatrix}, \tag{29}$$

where the well-known notation

$$\mathbf{F} = \mathbf{B}\mathbf{K}^{+}\mathbf{B}^{T}, \quad \mathbf{G} = -\mathbf{B}\mathbf{R}, \quad \mathbf{d} = \mathbf{B}\mathbf{K}^{+}\mathbf{f}, \quad \mathbf{e} = -\mathbf{R}^{T}\mathbf{f} \tag{30}$$

is used.

In reference [6], a constitutive law for the Lagrange multipliers and the discontinuity was introduced in the form

$$\mathbf{c} = \mathbf{H}\lambda, \tag{31}$$

where the compliance matrix, $\mathbf{H}$, was defined. The coarse problem can be rewritten to the form

$$\begin{pmatrix} \mathbf{F}+\mathbf{H} & \mathbf{G} \\ \mathbf{G}^{T} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \lambda \\ \alpha \end{pmatrix} = \begin{pmatrix} \mathbf{d} \\ \mathbf{e} \end{pmatrix}. \tag{32}$$

The system of equations (32) is solved by the modified preconditioned conjugate gradient method.

Comparison of (26) and (31) reveals the following equalities

$$\mathbf{c} = [\![\mathbf{u}]\!]_h = \mathbf{H}\lambda = \mathbf{K}_{int}^{-1}(\omega_h)\lambda_h. \tag{33}$$

4 Numerical Examples

The proposed strategy is applied to the end-notched flexure (ENF) test and the mixed-mode flexure (MMF) test used in reference [8]. The set up of the tests is depicted in Fig. 2. The material parameters are the following: Young's modulus of elasticity $E =$

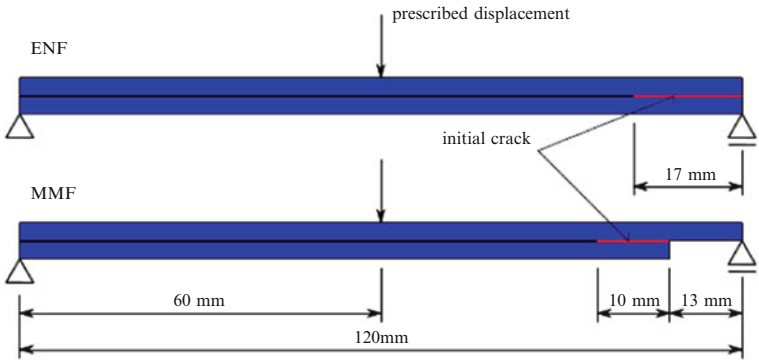


Fig. 2. End-notched flexure (ENF) and mixed-mode flexure (MMF) tests

75 GPa, Poisson's ratio $\nu = 0.3$, critical stress $\sigma_{max} = 3.602$ MPa, critical opening $\Delta = 0.011$ mm, fracture toughness $G = 0.02$ N/mm, mode mixity parameter $\beta = 0.472$. The structures are discretized by quadrilateral finite elements with bi-linear basis functions. They are loaded by prescribed displacements in the center.

The load-deflection curves for both tests are depicted in Figs. 3 and 4 Very good agreement with results published in [8] and [7] is obtained.

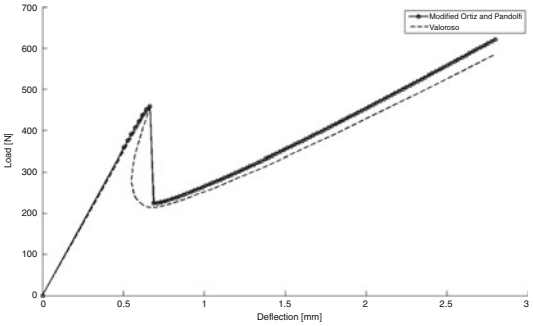


Fig. 3. Load-deflection curves for ENF test

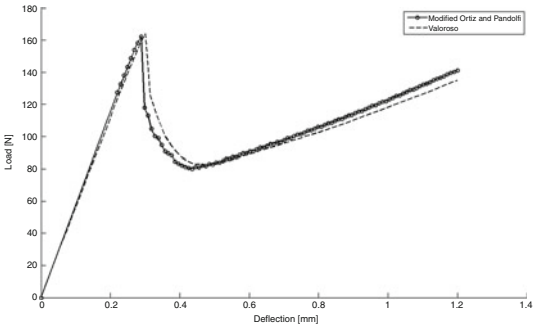


Fig. 4. Load-deflection curves for MMF test

5 Conclusions

Description of the imperfect material interface based on the compliance matrix **H** introduced in [6] was generalized with help of the energy-based delamination model described in [2]. This formulation uses piecewise constant approximation of damage variables and as such it allows to express the interfacial stiffness matrix easily.

Acknowledgments Financial support for this work was provided by project number VZ 03 CEZ MSM 6840770003 of the Ministry of Education of the Czech Republic. The financial support is gratefully acknowledged.

Bibliography

[1] Charbel Farhat and François-Xavier Roux. Implicit parallel processing in structural mechanics. *Comput. Mech. Adv.*, 2(1):124, 1994. ISSN 0927–7951.

[2] Pavel Gruber and Jan Zeman. A rate-independent model for composite materials with imperfect interfaces based on energy minimization. In B. H. V. Topping, J. M. Adam, F. J. Pallarés, R. Bru, and M. L. Romero, editors, *Proceedings of the Seventh International Conference on Engineering Computational Technology*, Stirlingshire, Scotland, 2010. Civil-Comp Press. paper 10.

[3] C.Y. Huia, A. Ruina, R. Long, and A. Jagota. Cohesive zone models and fracture. *The Journal of Adhesion*, 87:1–52, 2011.

[4] Jaroslav Kruis. Domain decomposition methods on parallel computers. In B. H. V. Topping and C. A. Mota Soares, editors, *Progress in Engineering Computational Technology*, pages 299–322. Saxe-Coburg Publications, Stirling, Scotland, UK, 2004.

[5] Jaroslav Kruis. *Computational Technology Reviews*, volume 3, chapter Domain Decomposition Methods in Engineering Computations. Saxe-Coburg Publications, Stirlingshire, Scotland, 2011.

[6] Jaroslav Kruis and Zdeněk Bittnar. Reinforcement-matrix interaction modeled by FETI method. In *Domain decomposition methods in science and engineering XVII*, volume 60 of *Lect. Notes Comput. Sci. Eng.*, pages 567–573. Springer, Berlin, 2008. doi: 10.1007/978-3-540-75199-1_71. URL http://dx.doi.org/10.1007/978-3-540-75199-1_71.

[7] M. Ortiz and A. Pandolfi. Finite-deformation irreversible cohesive elements for three-dimensional crack propagation analysis. *International Journal for Numerical Methods in Engineering*, 44:1267–1282, 1999.

[8] N. Valoroso and L. Champaney. A damage-mechanics-based approach for modelling decohesion in adhesively bonded assemblies. *Engineering Fracture Mechanics*, 73:2774–2801, 2006.

A Comparison of TFETI and TBETI for Numerical Solution of Engineering Problems of Contact Mechanics

D. Lukáš[2], M. Sadowská[1], T. Kozubek[1], A. Markopoulos[2], and T. Brzobohatý[2]

[1] FEECS, VŠB - TU Ostrava, Czech Republic, dalibor.lukas@vsb.cz, marie.sadowska@vsb.cz, tomas.kozubek@vsb.cz

[2] FME, VŠB - TU Ostrava, Czech Republic, alexandros.markopoulos@vsb.cz, tomas.brzobohaty@vsb.cz

Summary. Since the introduction of Finite Element Tearing and Interconnecting (FETI) by Farhat and Roux in 1991, the method has been recognized to be an efficient parallel technique for the solution of partial differential equations. In 2003 Langer and Steinbach formulated its boundary element counterpart (BETI), which reduces the problem dimension to subdomain boundaries. Recently, we have applied both FETI and BETI to contact problems of mechanics. In this paper we numerically compare their variants bearing the prefix Total (TFETI/TBETI) on a frictionless Hertz contact problem and on a realistic problem with a given friction.

1 Introduction

One of the leading representatives of domain decomposition methods is the Finite Element Tearing and Interconnecting (FETI) proposed by Farhat and Roux [8]. It relies on a finite element discretization of a linear elliptic boundary value problem and a nonoverlapping decomposition of the related geometric computational domain into subdomains. Resulting local subproblems are glued by means of Lagrange multipliers. The dual coarse problem is solved for the Lagrange multipliers by the method of conjugate gradients. Farhat et al. [9] proved that the condition number of the Schur complement, which arises from the elimination of the interior degrees of freedom, preconditioned by a projector orthogonal to the kernel is proportional to H/h, where H denotes the maximal subdomain diameter and h is the finite element discretization parameter. Moreover, [15] proved a polylogarithmic bound on the condition number of the Schur complement preconditioned by the Dirichlet preconditioner. This result was extended by Klawonn and Widlund [10] to the case of a redundant set of Lagrange multipliers and the correct (multiplicity or stiffness) scaling.

As the Lagrange multipliers live on the skeleton of the decomposition, it is very natural to employ a boundary integral representation of solutions to the local subproblems. This is the Boundary Element Tearing and Interconnecting (BETI) method, which was formulated and analyzed by Langer and Steinbach [13]. The

R. Bank et al. (eds.), *Domain Decomposition Methods in Science and Engineering XX*, Lecture Notes in Computational Science and Engineering 91,
DOI 10.1007/978-3-642-35275-1_40,

resulting discretized Steklov-Poincaré operators, which relate the local Cauchy data, are proved to be spectrally equivalent to the finite element Schur complements which eliminate interior degrees of freedom. An application of fully populated boundary element (BE) matrices can be sparsified to a linear complexity (up to a logarithmic factor), cf. [18]. Steinbach and Wendland [21] proposed a preconditioning of the BE matrices by related opposite order BE operators. The latter two accelaration techniques were exploited by Langer et al. [14] within the BETI method formulated in a twofold saddle-point system. It turned to be natural to impose additional Lagrange multipliers along the Dirichlet boundary, which was independently introduced as Total FETI (TFETI) by Dostál et al. [6] and as All-Floating BETI by Of [16], see also [17].

An extension of FETI and BETI methods to contact problems is a challenging task due to the strong nonlinearity of the variational inequality under consideration. To name a few of many research groups attacking this problem, see [1, 11, 20, 22]. The base for our development is a theoretically supported scalable algorithm for both coercive and semicoercive contact problems presented by Dostál et al. [7] and in the monograph by Dostál [5]. The first scalability results using TBETI for the scalar variational inequalities and the coercive contact problems were presented only recently by Bouchala et al. [2, 3], respectively. We also refer to [19].

The aim of this paper is to numerically compare TFETI and TBETI for two realistic problems. In Sect. 2 we recall the algebraic formulation of the TFETI and TBETI methods for contact problems. In Sect. 3 we describe different representations of the Schur complement. In Sect. 4 we compare the methods for the 3-dimensional (3d) Hertz contact problem without a friction and for a 3d contact problem of a ball bearing with a given friction. In Sect. 5 we conclude.

2 TFETI/TBETI Formulations

Both TFETI and TBETI methods for contact problems of mechanics lead, after a discretization, to the following problem:

$$\min_u \frac{1}{2}\langle Su,u\rangle - \langle f,u\rangle \text{ subject to } B_{\mathscr{I}}u \le c_{\mathscr{I}} \text{ and } B_{\mathscr{E}}u = c_{\mathscr{E}},$$

where we search for the local boundary displacement fields $u := (u_1,\ldots,u_p)$ with p being the number of subdomains. The Hessian $S := \mathrm{diag}(S_1,\ldots,S_p)$ consists of the Schur complements which are local Neumann finite element stiffness matrices eliminated to subdomain boundaries in the case of TFETI, and which are symmetric boundary element discretizations of local Steklov-Poincaré operators in the case of TBETI. Note that $\mathrm{Ker}\, S_i$ is the space spanned by six linearized local rigid body modes. In $f := (f_1,\ldots,f_p)$ we cummulate local boundary tractions. Further, $B_{\mathscr{E}}$ is a full rank sign matrix, the first part of which interconnects teared degrees of freedom with corresponding first part of $c_{\mathscr{E}}$ to be zero, while the second parts of $B_{\mathscr{E}}$ and $c_{\mathscr{E}}$ realize the Dirichlet boundary condition. Finally, the inequality with $B_{\mathscr{I}}$, $c_{\mathscr{I}}$ prescribes linearized non-penetration conditions.

Due to expensive projections onto the linear inequality constraints, we switch to the dual formulation with simple bound and equality constraints

$$\min_{\lambda_{\mathscr{I}}\geq 0} \frac{1}{2}\langle BS^{+}B^{T}\lambda,\lambda\rangle - \langle BS^{+}f-c,\lambda\rangle \text{ s.t. } (B^{T}\lambda-f)\perp \mathrm{Ker}\, S,$$

where we introduce Lagrange multipliers $\lambda := (\lambda_{\mathscr{I}},\lambda_{\mathscr{E}})$ with $\mathscr{I}$ and $\mathscr{E}$ referring to the inequality and equality constraints, respectively. Further, we cover $B_{\mathscr{I}}$, $B_{\mathscr{E}}$ by B and similarly $c := (c_{\mathscr{I}},c_{\mathscr{E}})$. Let S^{+} be a pseudoinverse of S, i.e., $SS^{+}g = g$ for any $g\perp \mathrm{Ker}\, S$. Let us denote by $R := \mathrm{diag}(R_1,\ldots,R_p)$ the column basis of $\mathrm{Ker}\, S$ consisting of local rigid body modes R_i and by P the orthogonal projector from $\mathrm{Im}\, B$ onto $\mathrm{Ker}\, R^{T}B^{T} = (\mathrm{Ker}\, S)^{\perp}$. To homogenize the linear (orthogonality) constraint, assume we are given a feasible λ_0 and search for $\lambda := \tilde{\lambda}+\lambda_0$. Returning to the old notation, we arrive at the following constrained quadratic programming problem preconditioned by the projector P and regularized by the complementary projector $Q := I-P$:

$$\min_{\lambda_{\mathscr{I}}\geq -(\lambda_0)_{\mathscr{I}}} \frac{1}{2}\left\langle \left(\frac{1}{\rho}PFP+Q\right)\lambda,\lambda\right\rangle - \left\langle \frac{1}{\rho}P(BS^{+}f_0-c),\lambda\right\rangle \text{ s.t. } R^{T}B^{T}\lambda = 0, \quad (1)$$

where $F := BS^{+}B^{T}$ and $f_0 := f-B^{T}\lambda_0$. Finally, we scale the cost function by $\rho \approx \|PFP\|$. Now from Theorem 3.2 of [9] and from the spectral equivalence of local boundary element and finite element Schur complements S_i, see Lemma 3.2 of [13], we have the following optimality result valid for both TFETI and TBETI.

Theorem 1. *Denote $\mathscr{H} := (1/\rho)PFP+Q$. There exist $c,C>0$ independent of h, H so that*

$$\lambda_{\min}(\mathscr{H}|\mathrm{Im}\, P) \geq c\frac{h}{H} \quad \textit{and} \quad \lambda_{\max}(\mathscr{H}|\mathrm{Im}\, P) = \|\mathscr{H}\| \leq C.$$

We are now in the position to use the augmented Lagrangian algorithm developed by Dostál [4], see also [5], for the solution of our constraint minimization problem (1). We mention that this algorithm is in some sense optimal.

3 Schur Complements

The local Schur complements S_i represent symmetric discretizations of the Steklov-Poincaré operator $\tilde{S}_i$ mapping the Dirichlet data to the Neumann data. In particular, $\tilde{S}_i(u_i) := \sigma_i(\varepsilon(\tilde{u}_i))\cdot n_i$ in the case of elastostatics, where n_i is the outward unit normal to the subdomain Ω_i, $\sigma_i(\varepsilon(\tilde{u}_i))$ denotes the elastostatic stress evaluated using the local linearized Hooke's law between the stress σ_i and the strain $\varepsilon(\tilde{u}_i)$, and where $\tilde{u}_i$ solves the following inhomogeneous Dirichlet boundary value problem:

$$\mathrm{div}\, \sigma_i(\varepsilon(\tilde{u}_i(x))) = 0 \text{ in } \Omega_i, \quad \tilde{u}_i(x) = u_i(x) \text{ on } \partial\Omega_i. \quad (2)$$

In the case of TFETI we solve (2) approximately by the finite element method. The approximation of $\tilde{S}_i$ is then as follows:

$$S_i := (A_i)_{BB} - (A_i)_{BI}(A_i)_{II}^{-1}(A_i)_{IB},$$

where $(A_i)_{jk} := \int_{\Omega_i} \sigma_i(\varepsilon(\varphi_j^{(i)}(x))) : \varepsilon(\varphi_k^{(i)}(x))\, dx$ is the Neumann finite element matrix assembled in the vector lowest order nodal basis functions $\varphi_j^{(i)}$, and where B and I are the sets of indices of boundary and interior degrees of freedom, respectively.

In the case of TBETI the interior degrees of freedom are already eliminated in the continuous formulation via a boundary integral representation of $\tilde{u}_i(x)$ while making use of the known elastostatic fundamental solution. After the lowest order Galerkin boundary element discretization, we arrive at the following relation between the approximated nodal based Dirichlet data, still denoted by u_i, and the element-based Neumann data, denoted by $t_i \approx \sigma_i(\varepsilon(\tilde{u}_i)) \cdot n_i$:

$$\begin{pmatrix} u_i \\ t_i \end{pmatrix} = \begin{pmatrix} (1/2)M_i - K_i & V_i \\ D_i & ((1/2)M_i + K_i)^T \end{pmatrix} \begin{pmatrix} u_i \\ t_i \end{pmatrix}$$

with fully populated boundary element matrices V_i, K_i, and D_i, which are referred to as single-layer, double-layer, and hypersingular matrix, respectively, and with the boundary mass matrix M_i. We then employ the following symmetric approximation of the Schur complement $\tilde{S}_i$:

$$S_i := D_i + ((1/2)M_i + K_i)^T V_i^{-1} ((1/2)M_i + K_i).$$

4 Numerical Comparison

All the presented simulations are performed using a parallel Matlab within our MatSol library, see [12]. The implementations of TFETI and TBETI are consistent. The only point where they differ is assembling of FEM and BEM matrices and subsequent Cholesky factorizations. In the preprocessing phase times for the BEM matrices assembling dominate. Our simulations were run on a cluster of 48 cores with 2.5 GHz and the infiband interface, which are equipped with licences of Matlab parallel computing engine.

First we consider a frictionless 3–dimensional Hertz problem, as depicted in Fig. 1, with the Young modulus $2.1 \cdot 10^5$ MPa and the Poisson ratio 0.3, where the ball is loaded from top by the force 5,000 N. ANSYS discretization of the two bodies is decomposed by METIS into 1,024 subdomains. The comparison of TFETI and TBETI in terms of computational times and number of Hessian multiplications is given in Table 1. In Fig. 2 we can see a fine correspondence of contact pressures computed by TFETI and TBETI to the analytical solution. The convergence criterion was the decay of the dual error to 10^{-6} relatively to the initial dual residuum.

In the second example we solve the contact problem of ball bearing, which consists of 10 bodies. We impose Dirichlet boundary condition along the outer perimeter and load the opposite part of the inner diameter with the force 4,500 N as depicted in Fig. 3. The Young modulus and the Poisson ratio of the balls and rings are

Fig. 1. Geometry of the Hertz problem

method	number of primal DOFs	number of dual DOFs	preprocessing time	solution time	number of Hessian applications
TFETI	4,088,832	926,435	21 min	1 h 49 min	593
TBETI	1,849,344	926,435	1h 33 min	1 h 30 min	667

Table 1. Numerical performance of TFETI and TBETI applied to the Hertz problem

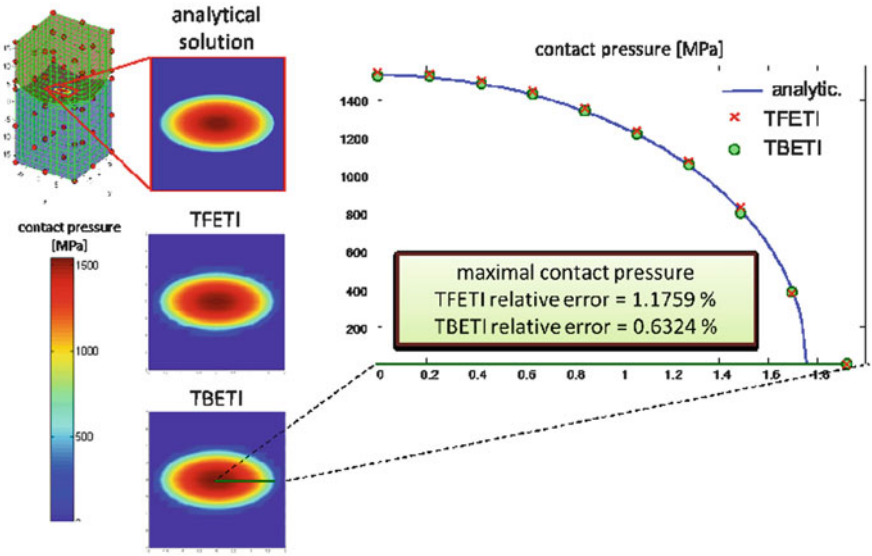


Fig. 2. Correspondence of numerical Hertz contact pressures to the analytic solution

$2.1 \cdot 10^5$ MPa and 0.3, respectively. Those of the cage are $2 \cdot 10^4$ MPa and 0.4, respectively. To get rid of the rigid body modes in the solution we introduce a small boundary gravitation term for each of the bodies. The discretized geometry was decomposed into 960 subdomains. Numerical comparison of TFETI and TBETI is shown in Table 2 and the resulting vertical displacement field is depicted in Fig. 4.

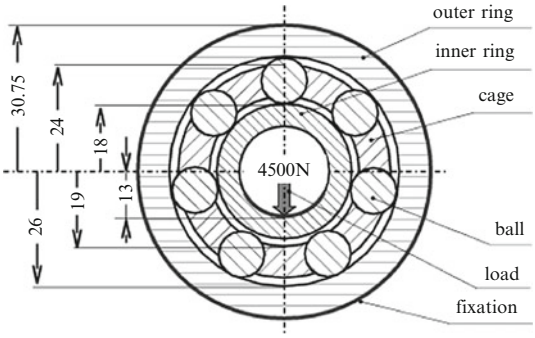


Fig. 3. Ball bearing: geometry, applied force and the Dirichlet boundary

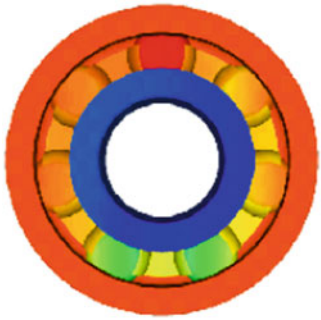


Fig. 4. Ball bearing: vertical component of the computed displacement field

method	number of primal DOFs	number of dual DOFs	preprocessing time	solution time	number of Hessian applications
TFETI	1,759,782	493,018	129 s	2 h 5 min	3203
TBETI	1,071,759	493,018	715 s	1 h 52 min	2757

Table 2. Numerical performance of TFETI and TBETI applied to the ball bearing problem

5 Conclusion

In the paper we compared TFETI and TBETI and numerically documented their performance for two engineering problems. Concerning timings and numbers of iterations it was shown that the methods are rather equal up to the assembling phase, which is more expensive in TBETI case. On the other hand, the accuracy of the boundary element discretization is usually much higher than the corresponding finite element discretization. This statement is supported by the theory provided that the solution is sufficiently regular. It can be also seen from Fig. 2, where one can guess that the TFETI relative error of 1.1759 % can be obtained with much less TBETI degrees of freedom.

Acknowledgments This research has been financially supported by the grants GA CR 201/07/0294 and the Ministry of Education of the Czech Republic No. MSM6198910027. This work was also supported by the European Regional Development Fund in the IT4Innovations Centre of Excellence project (CZ.1.05/1.1.00/02.0070).

Bibliography

[1] P. Avery and C. Farhat. The FETI family of domain decomposition methods for inequality-constrained quadratic programming: Application to contact problems with conforming and nonconforming interfaces. *Comput Method Appl M*, 198:1673–1683, 2009.

[2] J. Bouchala, Z. Dostál, and M. Sadowská. Theoretically supported scalable BETI method for variational inequaltities. *Computing*, 82:53–75, 2008.

[3] J. Bouchala, Z. Dostál, and M. Sadowská. Scalable Total BETI based algorithm for 3D coercive contact problems for linear elastostatics. *Computing*, 85:189–217, 2009.

[4] Z. Dostál. An optimal algorithm for bound and equality constrained quadratic programming problems with bounded spectrum. *Computing*, 78:311–328, 2006.

[5] Z. Dostál. *Optimal Quadratic Programming Algorithms with Applications to Variational Inequalities. 1dt edition. SOIA*, volume 23. Springer - Verlag, New York, 2009.

[6] Z. Dostál, D. Horák, and R. Kučera. Total FETI - an easier implementable variant of the FETI method for numerical solution of elliptic PDE. *Commun Numer Meth En*, 196:1155–1162, 2006.

[7] Z. Dostál, T. Kozubek, V. Vondrák, T. Brzobohatý, and A. Markopoulos. Scalable TFETI algorithm for the solution of multibody contact problems of elasticity. *Int J Numer Meth Engng*, 82:1384–1405, 2010.

[8] C. Farhat and F.-X. Roux. A method of finite element tearing and interconnecting and its parallel solution algorithm. *Int J Numer Meth Eng*, 32:1205–1227, 1991.

[9] C. Farhat, J. Mandel, and F.-X. Roux. Optimal convergence properties of the FETI domain decomposition method. *Comput Method Appl Mech Eng*, 115: 365–385, 1994.
[10] A. Klawonn and O.B. Widlund. FETI and Neumann-Neumann iterative substructuring methods: Connections and new results. *Comm Pure Appl Math*, 54: 57–90, 2001.
[11] R. Kornhuber and R. Krause. Adaptive multigrid methods for Signorini's problem in linear elasticity. *Comput Visual Sci*, 4:9–20, 2001.
[12] T. Kozubek, A. Markopoulos, T. Brzobohatý, R. Kučera, V. Vondrák, and Z. Dostál. MatSol – MATLAB efficient solvers for problems in engineering. http://www.am.vsb.cz/matsol.
[13] U. Langer and O. Steinbach. Boundary element tearing and interconnecting methods. *Computing*, 71:205–228, 2003.
[14] U. Langer, G. Of, O. Steinbach, and W. Zulehner. Inexact data-sparse boundary element tearing and interconnecting methods. *SIAM J Sci Comp*, 29:290–314, 2007.
[15] J. Mandel and R. Tezaur. Convergence of a substructuring method with Lagrange multipliers. *Numer Math*, 73:473–487, 1996.
[16] G. Of. The All-floating BETI method: numerical results. In *Domain Decomposition Methods in Science and Engineering XVII*, volume 60 of *LNCSE*, pages 295–302. Springer, Berlin, Heidelberg, 2008.
[17] G. Of and O. Steinbach. The all-floating boundary element tearing and interconnecting method. *J Numer Math*, 17:277–298, 2009.
[18] G. Of, O. Steinbach, and W.L. Wendland. Applications of a fast multipole Galerkin in boundary element method in linear elastostatics. *Comput Visual Sci*, 8:201–209, 2005.
[19] M. Sadowská, Z. Dostál, T. Kozubek, A. Markopoulos, and J. Bouchala. Scalable total BETI based solver for 3D multibody frictionless contact problems in mechanical engineering. *Eng Anal Bound Elem*, 35:330–341, 2011.
[20] J. Schöberl. Solving the Signorini problem on the basis of domain decomposition technique. *Computing*, 60:323–344, 1998.
[21] O. Steinbach and W.L. Wendland. The construction of some efficient preconditioners in the boundary element method. *Adv Comput Math*, 9:191–216, 1998.
[22] B.I. Wolmuth and R. Krause. Monotone methods on nonmatching grids for nonlinear contact problems. *SIAM J Sci Comput*, 25:324–347, 2003.

FETI-DP for Elasticity with Almost Incompressible Material Components

Sabrina Gippert, Axel Klawonn, and Oliver Rheinbach

Lehrstuhl für Numerische Mathematik, Fakultät für Mathematik, Universität Duisburg-Essen, D-45117 Essen, Germany. http://www.numerik.uni-duisburg-essen.de
{sabrina.gippert,axel.klawonn,oliver.rheinbach}@uni-duisburg-essen.de

1 Introduction

The purpose of this article is to present convergence bounds and some preliminary numerical results for a special category of problems of compressible and almost incompressible linear elasticity when using FETI-DP or BDDC domain decomposition methods.

We consider compressible and almost incompressible elasticity on the computational domain $\Omega \subset \mathbb{R}^3$ which is partitioned into a number of subdomains. We introduce nodes in the interior of the subdomains and on the interface. We distribute the material parameters such that in a neighborhood of the interface we have compressible and in the interior of a subdomain we have almost incompressible linear elasticity. Thus, each subdomain may contain an almost incompressible component in its interior surrounded by a hull of compressible material. We will also refer to this component as the incompressible inclusion.

By performing our analysis on the compressible hull, we can prove new condition number bounds. Such bounds will depend on the variation of the Poisson ratio ν in a neighborhood of the interface of the subdomains. More precisely, for compressible linear elasticity in a neighborhood of the interface and almost incompressible linear elasticity in the interior of the subdomains, we can prove a polylogarithmic condition number bound for the preconditioned FETI-DP system, which also depends on the thickness η of the compressible hull.

The condition number estimate presented in this contribution is based on the theory developed in [8] for compressible linear elasticity. It can be seen as an extension to certain configurations of incompressible components. For an algorithmic description of the FETI-DP method and the primal constraints applied in this paper, we refer to [5, 6]. The current work can also be seen as an extension of the work of [13–15]. There, the one-level FETI method for scalar elliptic problems is analyzed for special cases of coefficient jumps inside subdomains.

Coarse spaces for iterative substructuring methods that are robust either with respect to exact incompressibility constraints or with respect to almost incompressibility have been known for some time. For earlier work on Neumann-Neumann,

R. Bank et al. (eds.), *Domain Decomposition Methods in Science and Engineering XX*, Lecture Notes in Computational Science and Engineering 91,
DOI 10.1007/978-3-642-35275-1_41,

FETI-DP, and BDDC methods for (almost) incompressible elasticity, see, e.g., [4, 9, 10, 12].

2 Almost Incompressible Linear Elasticity

Let $\Omega \subset \mathbb{R}^3$ be a polytope, which can be decomposed into smaller cubic subdomains. We can allow also for subdomains that are images of cubes under a reasonable mapping.

The domain is fixed on $\partial\Omega_D \subset \partial\Omega$, i.e., we impose Dirichlet boundary conditions, and the remaining part $\partial\Omega_N = \partial\Omega \setminus \partial\Omega_D$ is subject to a surface force g. Let $H^1_0(\Omega, \partial\Omega_D) := \left\{ v \in (H^1(\Omega))^3 : \ v\,|_{\partial\Omega_D} = 0 \right\}$ be the Sobolev space which is appropriate for the variational formulation. Furthermore, the linearized strain tensor $\varepsilon = (\varepsilon_{ij})_{ij}$ is defined as $\varepsilon(u) = \frac{1}{2}(\nabla u + (\nabla u)^T)$ with $u \in (H^1(\Omega))^3$.

Then, the linear elasticity problem is defined as follows.
Find the displacement $u \in H^1_0(\Omega, \partial\Omega_D)$, such that for all $v \in H^1_0(\Omega, \partial\Omega_D)$

$$\int_\Omega G\, \varepsilon(u) : \varepsilon(v)\, dx + \int_\Omega G\, \beta\, \mathrm{div}(u)\, \mathrm{div}(v)\, dx = < F, v >$$

with the material parameters G, β, and the right hand side

$$< F, v > = \int_\Omega f^T v\, dx + \int_{\partial\Omega_N} g^T v\, d\sigma.$$

The material parameters G and β can also be expressed using Young's modulus E and the Poisson ratio ν by $G = \frac{E}{1+\nu}$ and $\beta = \frac{\nu}{1-2\nu}$. We analyze linear elasticity problems with different material components. For the compressible part we use the standard displacement formulation, i.e., we discretize the displacement by piecewise quadratic tetrahedral finite elements.

For almost incompressible linear elasticity, i.e., when $\nu \to \frac{1}{2}$, the value of β tends to infinity, and the discretization of the standard displacement formulation of linear elasticity by low order finite elements leads to locking effects and slow convergence. As a remedy the displacement problem is replaced by a mixed formulation. Therefore, we introduce the pressure $p := G\,\beta\ \mathrm{div}(u) \in L_2(\Omega)$ as an auxiliary variable.

We consider the problem: Find $(u, p) \in H^1_0(\Omega, \partial\Omega_D) \times L_2(\Omega)$, such that

$$\begin{aligned} \int_\Omega G\, \varepsilon(u) : \varepsilon(v)\, dx + \int_\Omega \mathrm{div}(v)\, p\, dx &= \langle F, v \rangle \quad \forall v \in H^1_0(\Omega, \partial\Omega_D) \\ \int_\Omega \mathrm{div}(u)\, q\, dx - \int_\Omega \frac{1}{G\,\beta}\, p\, q\, dx &= 0 \quad \forall q \in L_2(\Omega). \end{aligned}$$

It is well-known that in the case of almost incompressible linear elasticity, the solution of this mixed formulation exists and is unique.

For the discretization of this mixed problem we can in principle use any inf-sup stable mixed finite element method. For simplicity we use $Q_2 - P_0$ mixed finite elements, i.e., we discretize the displacement with piecewise triquadratic hexahedral

finite elements and the pressure with piecewise constant elements. This discretization is known to be inf-sup stable, which, in 3D, can be derived from the results in [11]. To obtain again a symmetric positive definite problem, the pressure is statically condensated element-by-element. We assume that a triangulation τ_h of Ω is given with shape regular finite elements, having a typical diameter h. Additionally, we assume that Ω can be represented exactly as a union of finite elements.

The domain Ω is now decomposed into N nonoverlapping subdomains Ω_i, $i = 1,\ldots,N$, with diameter H_i. The resulting interface is given by $\Gamma := \bigcup_{i\neq j} (\partial\Omega_i \cap \partial\Omega_j) \setminus \partial\Omega_D$. We assume matching finite element nodes on the neighboring subdomains across the interface Γ.

Then, for each subdomain we assemble the corresponding linear system

$$K^{(i)}u^{(i)} = f^{(i)}.$$

From the local linear systems, we obtain the FETI-DP saddle point problem, which is solved using a FETI-DP algorithm; see e.g., [1, 2, 5–8] for references on this algorithm. In this article we consider in particular the algorithm given in [5, 6, 8]; see the latter references for an algorithmic description of parallel FETI-DP methods using primal edge constraints and a transformation of basis. Here, in particular, we assume that all vertices are primal and all edge averages over all subdomain edges are the same across the interface Γ.

In our analysis, each of the N subdomains may contain an almost incompressible part, here also called an inclusion or a component, surrounded by a compressible hull. We will specify the definitions of a hull as follows.

Definition 1. *The hull of a subdomain Ω_i with width η is defined as*

$$\Omega_{i,\eta} := \{x \in \Omega_i : \operatorname{dist}(x, \partial\Omega_i) < \eta\}; \qquad \textit{see Fig. 1.}$$

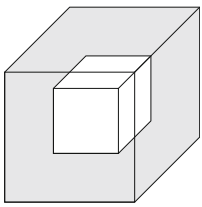


Fig. 1. $\Omega_{i,\eta}$: hull of Ω_i; see Definition 1

3 Convergence Analysis

In this section we provide a condition number estimate for the preconditioned FETI-DP matrix $M^{-1}F$, where F is the FETI-DP system matrix obtained from $K^{(i)}$ and

M^{-1} is the standard Dirichlet preconditioner; see [16]. We expand the convergence analysis, given in [8] for compressible linear elasticity, to the case where each subdomain can contain an almost incompressible inclusion surrounded by a compressible hull of thickness η. For the analysis, we make the following assumption; see [3] where the full details are provided.

Assumption 1 *For each subdomain, we have an inclusion which can be either almost incompressible or compressible, surrounded by a hull $\Omega_{i,\eta}$ of compressible material. The material coefficients $G(x)$ and $\beta(x)$ have a constant value in the interior inclusion and in the hull respectively, i.e.,*

$$G(x) = \begin{cases} G_{1,i} \; x \in \overline{\Omega}_{i,\eta} \\ G_{2,i} \; x \in \Omega_i \setminus \Omega_{i,\eta} \end{cases} \qquad \beta(x) = \begin{cases} \beta_{1,i} \; x \in \overline{\Omega}_{i,\eta} \\ \beta_{2,i} \; x \in \Omega_i \setminus \Omega_{i,\eta}. \end{cases}$$

Remark 1. Note that Assumption 1 allows that the Young modulus in the inclusion can be different from the one in the hull and that their quotient can be arbitrarily small or large.

The following assumption allows for the improved bound (2) in Theorem 1, which contains a linear factor H/η compared to the factor $(H/\eta)^4$ in (1).

Assumption 2 *For each subdomain Ω_i, $i = 1,\dots,N$, we assume that $G_{1,i} \leq k_i \cdot G_{2,i}$, where $k_i > 0$ is a constant independent of $h, H, \eta, G_{1,i}$, and $G_{2,i}$.*

In the analysis provided in [3], for the edge term estimate, we need a further assumption.

Assumption 3 *For any pair of subdomains (Ω_i, Ω_k) which have an edge in common, we assume that there exists an acceptable path $(\Omega_i, \Omega_{j_1}, \dots, \Omega_{j_n}, \Omega_k)$ from Ω_i to Ω_k, via a uniformly bounded number of other subdomains Ω_{i_q}, $q = 1,\dots n$, such that the coefficients G_{1,j_q} of the Ω_{i_q} satisfy the condition*

$$TOL \cdot G_{1,j_q} \geq \min(G_{1,i}, G_{1,k}), \; q = 1,\dots,n.$$

For a detailed description of the concept of acceptable paths, see [8, Sect. 5].

The following theorem is proven in [3].

Theorem 1. *Under the Assumptions 1 and 3, the condition number of the preconditioned FETI-DP system satisfies*

$$\kappa(M^{-1}F) \leq C \max(1, TOL) \left(1 + \log\left(\frac{H}{h}\right)\right) \left(1 + \log\left(\frac{\eta}{h}\right)\right) \left(\frac{H}{\eta}\right)^4, \qquad (1)$$

where $C > 0$ is independent of h, H, η, and the values of G_i and β_i, $i = 1,\dots,N$ and hence also of E_i and ν_i.

If additionally Assumption 2 is satisfied, we have

$$\kappa(M^{-1}F) \leq C \max(1, TOL) \left(1 + \log\left(\frac{H}{h}\right)\right)^2 \left(\frac{H}{\eta}\right), \qquad (2)$$

where $C > 0$ is independent of h, H, η, and the values of G_i and β_i, $i = 1,\dots,N$ and hence also of E_i and ν_i.

4 Numerical Results

In this section, we present our numerical results for a linear elasticity problem in three dimensions. We consider almost incompressible inclusions in the interior of the subdomains. The inclusions are always surrounded by a compressible hull with $\nu = 0.3$. We use a FETI-DP algorithm with vertices and edge averages as primal constraints to control the rigid body modes. For the algorithmic concept, see for example [8]. The numerical results confirm our theoretical estimates.

Our tests are divided into different categories.

4.1 Variable Thickness of the Compressible Hull

Here, we present results for $3 \times 3 \times 3$ subdomains, a fixed $H/h = 11$, and a fixed Poisson ratio $\nu = 0.499999$ in each inclusion and $\nu = 0.3$ in each hull. For these computations we vary the thickness of the hull, i.e., $\eta = 0, h, \ldots, 5h$; see Table 1. For the case $\eta = 0$, we obtain a large condition number of κ =1,597.8. This is not surprising since we use a coarse space designed for compressible linear elasticity. In this case using a different, larger coarse space in 3D is the remedy; see, e.g., [10] or [12].

It is striking that already a hull with a thickness of one element, i.e., $\eta = h$, is sufficient to obtain a good condition number which is then not improved significantly by further increasing η. As a result, the number of iteration steps does not change for $\eta = h, \ldots, 5h$. In our theory, see Theorem 1, for this configuration of coefficients, our bound is linear in H/η. From the numerical results in Table 1 we cannot conclude that the bound is sharp. This might be due to the fact, that in 3D we cannot choose our mesh fine enough. However, for 2D problems using very fine meshes the linear dependence on H/η can be observed numerically; see Table 2.

Table 1. Growing η; $H/h = 11$; $1/H = 3$.

η	iterations	condition number
0	50	1597.8
$1h$	32	12.366
$2h$	32	12.250
$3h$	32	12.230
$4h$	32	12.231
$5h$	32	12.233

Growing η for $3 \times 3 \times 3$ subdomains, $E = 210$ on the whole domain, $\nu = 0.499999$ in each inclusion, and $\nu = 0.3$ in each hull. The results show only a weak dependence on η.

Table 2. Growing η; 2D; $H/h = 200$; $1/H = 3$

η	iterations	condition number
1/100	47	199.906
2/100	41	102.081
3/100	42	70.719
4/100	36	54.674

Linear elasticity in 2D with $\Omega = [0,1]^2$, discretized with $Q_1 - P_0$ stabilized finite elements; for a description of the discretization, see, e.g., [9]. The domain is decomposed into square subdomains with sidelength H, having square inclusions and a hull of thickness η. The Poisson ratio in each inclusion is chosen as $\nu = 0.4999999$ and in each hull as $\nu = 0.3$. The Young modulus is chosen as $E = 1$ on the whole domain. The results confirm the linear dependence on H/η.

4.2 Variable Incompressibility in the Inclusions

In Table 3, we vary the Poisson ratio in the inclusions from $\nu = 0.4$ up to $\nu = 0.499999$ while choosing a fixed number of elements in each subdomain, i.e., $H/h = 7$, and a thickness of the hull of $\eta = h$. We see that the condition number is indeed bounded independently of the almost incompressibility in the inclusions as expected from Theorem 1.

Table 3. Growing ν; $H/h = 7$; $1/H = 3$; $\eta = h$.

ν	iterations	condition number
0.4	27	9.4841
0.49	28	9.5038
0.499	28	9.5063
0.4999	28	9.5049
0.49999	28	9.5066
0.499999	29	9.5066

Growing ν for $3 \times 3 \times 3$ subdomains, $\eta = h$, $\nu = 0.3$ in the hulls, and $E = 210$ on the whole domain. A hull with a thickness of one element is clearly sufficient to obtain a good condition number.

4.3 Variable Young's Modulus in the Inclusions Combined with Variable Incompressibility in the Inclusions

In a last set of experiments, see Table 4, we consider subdomains with inclusions of a high and low Young modulus, i.e., $E = 1e+4$ and $E = 1e-4$, either combined with a Poisson ratio of $\nu = 0.4$ or $\nu = 0.499999$; see Fig. 2. The Young modulus of

the hull is always $E = 1$ and its Poisson ratio is always $\nu = 0.3$. The four different parameter settings are determined by the number of the subdomain modulo four; see Fig. 2. In our theory, the condition number bound for such a configuration contains a factor $(H/\eta)^4$. However, the results in Table 4 are not worse than in the configurations where bound (1) of Theorem 1 applies, which contains only a linear H/η. The condition number is surprisingly low even if the thickness of the hull is only $\eta = h$. While this is a favorable result it also means that it is difficult to confirm numerically whether our theoretical bounds are sharp with respect to η.

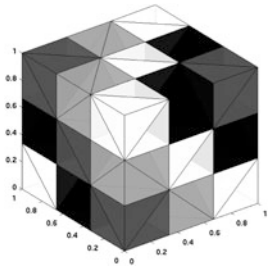


Fig. 2. Types of subdomains, see Table 4, identified by color

Table 4. Growing η; $H/h = 7$; $1/H = 3$.

distance η	iterations	condition number
0	> 250	13426
$1h$	36	11.956
$2h$	29	9.2575
$3h$	29	9.4767
$4h$	27	9.4812

Growing η for $3 \times 3 \times 3$ subdomains. Four different kind of material parameter settings in the inclusions: $E = 1e+4$ and $\nu = 0.4$; $E = 1e-4$ and $\nu = 0.4$; $E = 1e+4$ and $\nu = 0.499999$; $E = 1e-4$ and $\nu = 0.499999$; for all hulls: $E = 1, \nu = 0.3$.

Bibliography

[1] Charbel Farhat, Michel Lesoinne, and Kendall Pierson. A scalable dual-primal domain decomposition method. Preconditioning techniques for large sparse matrix problems in industrial applications. *Numer. Linear Algebra Appl.*, 7 (7–8):687–714, 2000.

[2] Charbel Farhat, Michel Lesoinne, Patrick LeTallec, Kendall Pierson, and Daniel Rixen. FETI-DP: a dual-primal unified FETI method. I. A faster alternative to the two-level FETI method. *Internat. J. Numer. Methods Engrg.*, 50(7):1523–1544, 2001.

[3] Sabrina Gippert, Axel Klawonn, and Oliver Rheinbach. Analysis of FETI-DP and BDDC for linear elasticity in 3D with almost incompressible components and varying coefficients inside subdomains. Submitted for publication June 2011, revised January 2012, 2012.

[4] Paulo Goldfeld, Luca F. Pavarino, and Olof B. Widlund. Balancing Neumann-Neumann preconditioners for mixed approximations of heterogeneous problems in linear elasticity. *Numer. Math.*, 95(2):283–324, 2003.

[5] Axel Klawonn and Oliver Rheinbach. A parallel implementation of Dual-Primal FETI methods for three dimensional linear elasticity using a transformation of basis. *SIAM J. Sci. Comput.*, 28:1886–1906, 2006.

[6] Axel Klawonn and Oliver Rheinbach. Robust FETI-DP methods for heterogeneous three dimensional linear elasticity problems. *Comput. Methods Appl. Mech. Engrg.*, 196:1400–1414, 2007.

[7] Axel Klawonn and Olof B. Widlund. Selecting Constraints in Dual-Primal FETI Methods for Elasticity in Three Dimensions. In *Domain Decomposition Methods in Science and Engineering, Lecture Notes Comput. Sci. Eng.*, volume 40, pages 67–81, 2005. Proceedings of the 15th International Conference on Domain Decomposition Methods, Berlin, July 21–25, 2003.

[8] Axel Klawonn and Olof B. Widlund. Dual-Primal FETI Methods for Linear Elasticity. *Comm. Pure Appl. Math*, 59:1523–1572, 2006.

[9] Axel Klawonn, Oliver Rheinbach, and Barbara I. Wohlmuth. Dual-primal iterative substructuring for almost incompressible elasticity. In David E. Keyes and Olof B. Widlund, editors, *Domain Decomposition Methods in Science and Engineering*, volume 55, pages 397–404. Springer-Verlag, Lecture Notes in Computational Science and Engineering, 2007. Proceedings of the 16th International Conference on Domain Decomposition Methods, New York, NY, January 12–15, 2005.

[10] Jing Li and Olof B. Widlund. BDDC algorithms for incompressible Stokes equations. *SIAM J. Numer. Anal.*, 44(6):2432–2455 (electronic), 2006. ISSN 0036–1429.

[11] Gunar Matthies and Lutz Tobiska. The inf-sup condition for the mapped $Q_k - P_{k-1}^{disc}$ element in arbitrary space dimensions. *Computing*, 69:119–139, 2002.

[12] Luca F. Pavarino, Olof B. Widlund, and Stefano Zampini. BDDC preconditioners for spectral element discretizations of almost incompressible elasticity in three dimension. *SIAM J. Sci. Comput.*, 32(6):3604–3626, 2010.

[13] Clemens Pechstein and Robert Scheichl. Analysis of FETI methods for multiscale PDEs. *Numer. Math.*, 111:293–333, 2008.

[14] Clemens Pechstein and Robert Scheichl. Scaling up through domain decomposition. *Applicable Analysis*, 88(10–11):1589–1608, 2009.

[15] Clemens Pechstein and Robert Scheichl. Analysis of FETI methods for multiscale PDEs- PartII: Interface variation. *Numer. Math.*, 118(3):485–529, 2011.

[16] Andrea Toselli and Olof B. Widlund. *Domain Decomposition Methods- Algorithms and Theory*, volume 34. Springer Series in Computational Mathematics, 2005.

An Alternative Coarse Space Method for Overlapping Schwarz Preconditioners for Raviart-Thomas Vector Fields

Duk-Soon Oh

Louisiana State University, 216 Johnston Hall, Baton Rouge, LA 70809, USA
duksoon@cims.nyu.edu

Summary. The purpose of this paper is to introduce an overlapping Schwarz method for vector field problems discretized with the lowest order Raviart-Thomas finite elements. The coarse component of the preconditioner is based on energy-minimizing discrete harmonic extensions and the local components consist of traditional solvers on overlapping subdomains. The approach has a couple of benefits compared to the previous methods. The algorithm can be implemented in an algebraic manner. Moreover, the method leads to a condition number independent of the values and jumps of the coefficients across the interface between the substructures. Supporting numerical examples to demonstrate the effectiveness are also presented.

1 Introduction

Domain decomposition methods can be categorized in two classes: overlapping Schwarz methods with overlapping subdomains and iterative substructuring methods with nonoverlapping subdomains. In this paper, we consider two level overlapping Schwarz algorithms. Such methods were originally developed for scalar elliptic problems; see [11, 15] and references therein. Later these methods have also been considered for solving vector fields problems posed in $H(\mathrm{div})$ and $H(\mathbf{curl})$; see [1, 9, 13]. Other types of algorithms, such as multigrid methods, classical iterative substructuring methods, balancing Neumann-Neumann, and FETI methods, have also been suggested in [3, 8, 12, 14, 16, 17]. Many nonoverlapping methods have been studied for discontinuous coefficients cases for vector fields problems. However, only few methods were introduced for the overlapping Schwarz methods in case of coefficients which have jumps.

In the domain decomposition theory, methods can often provide good scalability, i.e., the condition number of the preconditioned system will depend only on the size of the subdomain problems and not on any other parameters, e.g., the number of subdomains and jumps of the coefficients. For the purpose of handling the discontinuity, we borrow the advanced coarse space techniques of [6, 7] based on discrete harmonic extensions of coarse trace spaces developed for almost incompressible elasticity.

R. Bank et al. (eds.), *Domain Decomposition Methods in Science and Engineering XX*, Lecture Notes in Computational Science and Engineering 91,
DOI 10.1007/978-3-642-35275-1_42,

The rest part of this paper is organized as follows. We introduce a model problem and its finite element approximation in Sect. 2. In Sects. 3 and 4, we recall the overlapping Schwarz method and we suggest the alternative coarse algorithm, respectively. We next present the numerical results in Sect. 5. Finally, the conclusion of this paper is given in Sect. 6.

2 Discretized Problem

We consider the following second order partial differential equation for vector field problem posed in $H(\mathrm{div})$ in a bounded polyhedral domain Ω with a homogeneous boundary condition:

$$\begin{aligned} L\mathbf{u} := -\mathbf{grad}\,(\alpha \,\mathrm{div}\,\mathbf{u}) + \beta\,\mathbf{u} &= \mathbf{f} \text{ in } \Omega, \\ \mathbf{u}\cdot\mathbf{n} &= 0 \text{ on } \partial\Omega. \end{aligned} \tag{1}$$

Here we have positive coefficients $\alpha, \beta \in L^\infty(\Omega)$ and assume that $\mathbf{f}$ is in $(L^2(\Omega))^3$. The main focus of our work is on the coefficients α and β which have jumps across between the substructures.

The model problem (1) has many important applications, such as a mixed and least-squares formulation of certain types of second order partial differential equations [5, 17]. There are other types of applications related to $H(\mathrm{div})$, e.g., iterative solvers for the Reissner-Mindlin plate and the sequential regularization method for the Navier-Stokes equations. For more detail, see [2, 10].

We next consider a variational formulation of (1):

$$\mathbf{a}(\mathbf{u},\mathbf{v}) := \int_\Omega \alpha\,\mathrm{div}\,\mathbf{u}\,\mathrm{div}\,\mathbf{v}\,dx + \beta\,\mathbf{u}\cdot\mathbf{v}\,dx = \int_\Omega \mathbf{f}\cdot\mathbf{v}\,dx,\ \mathbf{v}\in H_0(\mathrm{div};\Omega). \tag{2}$$

We consider the lowest order Raviart-Thomas elements, conforming in $H(\mathrm{div})$, to obtain a discretized problem; see [4, Chap. 3]. We note that the degrees of freedom of the Raviart-Thomas elements are defined by the average values of the normal components over the faces.

Let us consider the variational problem (2). Restricting to the finite element space of the lowest order Raviart-Thomas elements with shape regular and quasi-uniform meshes, we obtain the following linear system:

$$Au = f, \tag{3}$$

where the matrix A is a stiffness matrix, u is a vector of degrees of freedom, and f is a known vector obtained from $\mathbf{f}$. We note that A is symmetric and positive definite.

3 Overlapping Schwarz Preconditioner

We consider a decomposition of the domain Ω into N nonoverlapping subdomains Ω_i, $i = 1,\cdots,N$. We next introduce extended subregions Ω_i' obtained from Ω_i by adding layers of elements and the interface Γ which is given by

$$\Gamma = \left(\bigcup_{i=0}^{N} \partial\Omega_i \right) \setminus \partial\Omega.$$

We consider a two-level overlapping Schwarz algorithm to solve the linear system (3). An overlapping Schwarz preconditioner usually has the following form:

$$P^{-1} = R_0^T A_0^{-1} R_0 + \sum_{i=1}^{N} R_i^T A_i^{-1} R_i, \tag{4}$$

where A_0 is the matrix of the global coarse problem, the A_i's are obtained from local subproblems related to the extended subdomains Ω_i', and R_0 and R_i's are restriction operators to the coarse space and local spaces, respectively; see [11, 15] for more details.

In [9, 13], model problems were designed for constant coefficients and convex domains to analyze the methods. In our work, we use more general assumptions: convex subdomains and coefficients which have jumps across the interface Γ.

In order to deal with this situation, we consider an alternative coarse space approach instead of traditional coarse interpolations. The basis functions for the new algorithm are based on energy-minimizing discrete harmonic extensions with given interface values. We use the corresponding discrete harmonic extensions of the boundary values of standard basis functions to construct new basis functions. We remark that this process can be performed locally and in parallel due to the fact that the basis functions are supported in just two subdomains. We also note that we do not need any coarse triangulation and this work can be done algebraically. With new alternative basis functions, we obtain the operator R_0 which defines the new basis and the matrix $A_0 = R_0 A R_0^T$ associated with the global coarse problem.

For the local components, we follow the traditional way. Each R_i is a rectangular matrix with elements equal to 0 and 1 and provides the indices relevant to an individual extended subdomain Ω_i'. Each $A_i = R_i A R_i^T$ is just the principal minor of the original stiffness matrix A defined by R_i. By using these matrices, we can build the local component $\sum_{i=1}^{N} R_i^T A_i^{-1} R_i$ of the Schwarz preconditioner.

4 The Coarse Component

In this section, we explain our approach in detail. We focus on the restriction operator R_0 onto the coarse space. Before we consider the alternative method, we introduce the conventional method in [9, 13]. The restriction operator is obtained by the interpolation from the subspaces defining the coarse component to the global space. More precisely, R_0 are exactly the coefficients obtained by interpolating the traditional coarse basis functions onto the fine mesh. We note that we need geometric information, e.g., coordinate information, to construct R_0.

Instead of the conventional coarse basis, we will use discrete harmonic extensions to define the new coarse basis functions. We first consider two adjacent subdomains Ω_i and Ω_j. We then have a coarse face $F_{ij} = \partial\Omega_i \cap \partial\Omega_j$. We note that each

coarse degree of freedom of our coarse component is related to each coarse face. Let u denote the vector of degrees of freedom for the original problem. Similarly, we consider the vectors of degrees of freedom $u_I^{(i)}$, $u_I^{(j)}$, and $u_{F_{ij}}$ associated with $\Omega_i \backslash \Gamma$, $\Omega_j \backslash \Gamma$, and F_{ij}, respectively. We then have restriction matrices $R_I^{(i)}$, $R_I^{(j)}$, and $R_{F_{ij}}$, i.e., $u_I^{(i)} = R_I^{(i)} u$, $u_I^{(j)} = R_I^{(j)} u$, and $u_{F_{ij}} = R_{F_{ij}} u$. We note that each restriction matrix has only one nonzero entry of unity per each row. We next introduce a submatrix of the stiffness matrix A. It corresponds to the two subdomains which have F_{ij} in common:

$$\begin{bmatrix} A_{II}^{(i)} & 0 & A_{IF_{ij}}^{(i)} \\ 0 & A_{II}^{(j)} & A_{IF_{ij}}^{(j)} \\ A_{F_{ij}I}^{(i)} & A_{F_{ij}I}^{(j)} & A_{F_{ij}F_{ij}} \end{bmatrix}.$$

We choose $u_{F_{ij}}^T = [1, 1, \cdots, 1]$ and introduce the local subproblems $A_{II}^{(i)} u_I^{(i)} + A_{IF_{ij}}^{(i)} u_{F_{ij}} = 0$ and $A_{II}^{(j)} u_I^{(j)} + A_{IF_{ij}}^{(j)} u_{F_{ij}} = 0$ to consider discrete harmonic extensions; see [15, Chap. 4.4]. Then, $u_I^{(i)}$ and $u_I^{(j)}$ are completely determined by $u_{F_{ij}}$, i.e., $u_I^{(i)} = E_i u_{F_{ij}}$ and $u_I^{(j)} = E_j u_{F_{ij}}$, where $E_i := -{A_{II}^{(i)}}^{-1} A_{IF_{ij}}^{(i)}$ and $E_j := -{A_{II}^{(j)}}^{-1} A_{IF_{ij}}^{(j)}$. We then obtain a coarse basis $u_{ij} = {R_I^{(i)}}^T u_I^{(i)} + {R_I^{(j)}}^T u_I^{(j)} + R_{F_{ij}}^T u_{F_{ij}}$ corresponding to F_{ij}. We can then construct the following form of our coarse interpolation matrix R_0 after the similar process:

$$R_0 := \begin{bmatrix} \vdots \\ - \; u_{ij}^T \; - \\ \vdots \end{bmatrix}.$$

As we mentioned earlier, we can obtain the coarse matrix A_0 by the Galerkin product $R_0 A R_0^T$. We remark that our alternative approach can be implemented in an algebraic manner and in parallel. However, we need to solve additional local Dirichlet-type subproblems to construct the coarse component compared to the conventional methods.

5 Numerical Experiments

We apply the overlapping Schwarz method with the energy-minimizing coarse space to our model problem. We use $\Omega = (0,1) \times (0,1) \times (0,1)$ and the lowest order hexahedral Raviart-Thomas elements. We decompose the domain into $N \times N \times N$ identical subdomains. In each subdomain, we assume that the coefficients α and β are constant. We consider cases where the coefficients have jumps across the interface between the subdomains, in particular, a checkerboard distribution pattern. Each subdomain Ω_i has side length $H = 1/N$ and each mesh cube has h as a minimum side length. We also introduce extended subdomains whose boundaries do not cut any

mesh elements with an overlap parameter δ between subdomains. We use the preconditioned conjugate gradient method to solve the preconditioned linear system

$$P^{-1}Au = P^{-1}f. \tag{5}$$

We stop the iteration when the residual l_2-norm has been reduced by a factor of 10^{-6}.

We perform two different kinds of experiments. We first fix the overlap parameter H/δ and vary H/h. We next fix the size of H/h and use various size of H/δ. We report the condition numbers estimated by the conjugate gradient method and the number of iterations. Tables 1 and 3 show the first results and Tables 2 and 4 show the results of the second experiments.

In the first set of experiments, we see that the condition numbers and the iteration counts do not depend on the size of H/h. In the second set, we can conclude that the condition numbers grow linearly with H/δ. For both cases, the condition numbers and iteration counts are quite independent of coefficients and the jumps of coefficients between the subdomains.

Table 1. Condition numbers and iteration counts. $\alpha_i = 1$ or specified values as indicated in a checkerboard pattern, $\beta_i \equiv 1$, $\frac{H}{\delta} = 4$, $H = \frac{1}{3}$, and $h = \frac{1}{12}, \frac{1}{24}, \frac{1}{48}$

	$\alpha_i = 0.01$		$\alpha_i = 0.1$		$\alpha_i = 1$		$\alpha_i = 10$		$\alpha_i = 100$	
$\frac{H}{h}$	cond	iters	cond	iters	cond	iters	cond	iters	cond	iters
4	8.23	15	8.90	16	9.16	17	8.92	16	8.25	15
8	8.39	16	9.01	17	9.20	18	9.00	17	8.28	16
16	8.23	16	8.99	17	9.22	19	8.98	17	8.28	16

Table 2. Condition numbers and iteration counts. $\alpha_i = 1$ or specified values as indicated in a checkerboard pattern, $\beta_i \equiv 1$, $\frac{H}{h} = 16$, $H = \frac{1}{3}$, and $h = \frac{1}{48}$

	$\alpha_i = 0.01$		$\alpha_i = 0.1$		$\alpha_i = 1$		$\alpha_i = 10$		$\alpha_i = 100$	
$\frac{H}{\delta}$	cond	iters	cond	iters	cond	iters	cond	iters	cond	iters
4	8.23	16	8.99	17	9.22	19	8.98	17	8.28	16
8	10.86	16	13.27	18	14.06	22	14.16	18	14.10	16
16	16.22	18	22.94	22	25.03	24	25.30	22	25.32	20

6 Conclusion

An alternative coarse space technique based on energy-minimizing discrete harmonic extensions for overlapping Schwarz algorithm for vector field problems posed in

Table 3. Condition numbers and iteration counts. $\beta_i = 1$ or specified values as indicated in a checkerboard pattern, $\alpha_i \equiv 1$, $\frac{H}{\delta} = 4$, $H = \frac{1}{3}$, and $h = \frac{1}{12}, \frac{1}{24}, \frac{1}{48}$

	$\beta_i = 0.01$		$\beta_i = 0.1$		$\beta_i = 1$		$\beta_i = 10$		$\beta_i = 100$	
$\frac{H}{h}$	cond	iters	cond	iters	cond	iters	cond	iters	cond	iters
4	8.18	15	8.36	16	9.16	17	8.68	17	8.36	16
8	8.18	17	8.46	18	9.20	18	8.65	18	8.37	18
16	8.18	17	8.45	18	9.22	19	8.62	18	8.37	18

Table 4. Condition numbers and iteration counts. $\beta_i = 1$ or specified values as indicated in a checkerboard pattern, $\alpha_i \equiv 1$, $\frac{H}{h} = 16$, $H = \frac{1}{3}$, and $h = \frac{1}{48}$

	$\beta_i = 0.01$		$\beta_i = 0.1$		$\beta_i = 1$		$\beta_i = 10$		$\beta_i = 100$	
$\frac{H}{\delta}$	cond	iters	cond	iters	cond	iters	cond	iters	cond	iters
4	8.18	17	8.45	18	9.22	19	8.62	18	8.37	18
8	8.50	17	9.98	18	14.06	22	13.48	21	9.43	19
16	9.34	17	13.13	21	25.03	24	24.79	22	12.56	19

$H(\mathrm{div})$ has been introduced and implemented. The numerical results show the usefulness of our method even in the presence of jumps of the coefficients between the substructures.

Acknowledgments The author would like to thank Prof. Olof Widlund for his suggestions and assistance. The author is also grateful to Dr. Clark Dohrmann for his useful comments. The work of the author has been supported by the NSF under Grant DMS-0914954.

Bibliography

[1] Douglas N. Arnold, Richard S. Falk, and R. Winther. Preconditioning in $H(\mathrm{div})$ and applications. *Math. Comp.*, 66(219):957–984, 1997.

[2] Douglas N. Arnold, Richard S. Falk, and Ragnar Winther. Preconditioning discrete approximations of the Reissner-Mindlin plate model. *RAIRO Modél. Math. Anal. Numér.*, 31(4):517–557, 1997.

[3] Douglas N. Arnold, Richard S. Falk, and Ragnar Winther. Multigrid in $H(\mathrm{div})$ and $H(\mathrm{curl})$. *Numer. Math.*, 85(2):197–217, 2000.

[4] Franco Brezzi and Michel Fortin. *Mixed and hybrid finite element methods*, volume 15 of *Springer Series in Computational Mathematics*. Springer-Verlag, New York, 1991.

[5] Z. Cai, R. Lazarov, T. A. Manteuffel, and S. F. McCormick. First-order system least squares for second-order partial differential equations. I. *SIAM J. Numer. Anal.*, 31(6):1785–1799, 1994.

[6] Clark R. Dohrmann and Olof B. Widlund. An overlapping Schwarz algorithm for almost incompressible elasticity. *SIAM J. Numer. Anal.*, 47(4):2897–2923, 2009.

[7] Clark R. Dohrmann and Olof B. Widlund. Hybrid domain decomposition algorithms for compressible and almost incompressible elasticity. *Internat. J. Numer. Methods Engrg.*, 82(2):157–183, 2010.

[8] R. Hiptmair. Multigrid method for Maxwell's equations. *SIAM J. Numer. Anal.*, 36(1):204–225 (electronic), 1999.

[9] Ralf Hiptmair and Andrea Toselli. Overlapping and multilevel Schwarz methods for vector valued elliptic problems in three dimensions. In *Parallel solution of partial differential equations (Minneapolis, MN, 1997)*, volume 120 of *IMA Vol. Math. Appl.*, pages 181–208. Springer, New York, 2000.

[10] Ping Lin. A sequential regularization method for time-dependent incompressible Navier-Stokes equations. *SIAM J. Numer. Anal.*, 34(3):1051–1071, 1997.

[11] Barry F. Smith, Petter E. Bjørstad, and William D. Gropp. *Domain decomposition*. Cambridge University Press, Cambridge, 1996. Parallel multilevel methods for elliptic partial differential equations.

[12] Andrea Toselli. Neumann-Neumann methods for vector field problems. *Electron. Trans. Numer. Anal.*, 11:1–24, 2000.

[13] Andrea Toselli. Overlapping Schwarz methods for Maxwell's equations in three dimensions. *Numer. Math.*, 86(4):733–752, 2000.

[14] Andrea Toselli and Axel Klawonn. A FETI domain decomposition method for edge element approximations in two dimensions with discontinuous coefficients. *SIAM J. Numer. Anal.*, 39(3):932–956 (electronic), 2001.

[15] Andrea Toselli and Olof Widlund. *Domain decomposition methods—algorithms and theory*, volume 34 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin, 2005.

[16] Barbara I. Wohlmuth. *Discretization methods and iterative solvers based on domain decomposition*, volume 17 of *Lecture Notes in Computational Science and Engineering*. Springer-Verlag, Berlin, 2001.

[17] Barbara I. Wohlmuth, Andrea Toselli, and Olof B. Widlund. An iterative substructuring method for Raviart-Thomas vector fields in three dimensions. *SIAM J. Numer. Anal.*, 37(5):1657–1676 (electronic), 2000.

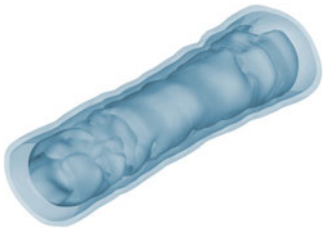
A Simultaneous Augmented Lagrange Approach for the Simulation of Soft Biological Tissue

Dirk Böse[1], Sarah Brinkhues[2], Raimund Erbel[1], Axel Klawonn[3], Oliver Rheinbach[3], and Jörg Schröder[2]

[1] Westdeutsches Herzzentrum, Universitätsklinikum Essen
[2] Fakultät für Ingenieurwissenschaften, Abteilung Bauwissenschaften, Institut für Mechanik, Universität Duisburg-Essen
sabrina.gippert@uni-koeln.de
[3] Mathematisches Institut, Universität zu Koeln, Weyertal 86-90, 50931 Koeln. Germany.
{axel.klawonn, oliver.rheinbach}@uni-koeln.de

Summary. In this paper, we consider the elastic deformation of arterial walls as occurring, e.g., in the process of a balloon angioplasty, a common treatment in the case of atherosclerosis. Soft biological tissue is an almost incompressible material. To account for this property in finite element simulations commonly used free energy functions contain terms penalizing volumetric changes. The incorporation of such penalty terms can, unfortunately, spoil the convergence of the nonlinear iteration scheme, i.e., of Newton's method, as well as of iterative solvers applied for the solution of the linearized systems of equations. We show that the augmented Lagrange method can improve the convergence of the linear and nonlinear iteration schemes while, at the same time, implementing a guaranteed bound for the volumetric change. Our finite element model of an atherosclerotic arterial segment, see Fig. 1, is constructed from intravascular ultrasound images; for details see [4].

Fig. 1. Finite element model of an atherosclerotic arterial segment 1.3*M* unknowns



R. Bank et al. (eds.), *Domain Decomposition Methods in Science and Engineering XX*, Lecture Notes in Computational Science and Engineering 91,
DOI 10.1007/978-3-642-35275-1_43,

1 Nonlinear Model and Algorithm

Biological tissues, such as arteries, are fiber enforced materials composed of an almost incompressible matrix substance with embedded collagen fibers. The arrangement of the fibers in arterial walls is characterized by two preferred directions helically wound along the artery. The material behavior of the collagen fiber bundles is represented by the superposition of two transversely isotropic models; see [12]. Thus, the strain energies are given by

$$\psi = \psi^{\mathrm{iso}}(\mathbf{C}) + \psi^{\mathrm{ti},(1)}(\mathbf{C},\mathbf{M}^{(1)}) + \psi^{\mathrm{ti},(2)}(\mathbf{C},\mathbf{M}^{(2)}). \tag{1}$$

Here, $\mathbf{F} := \nabla\varphi$ is the deformation gradient, $\mathbf{C} := \mathbf{F}^T\mathbf{F}$ the right Cauchy-Green-tensor, and $\mathbf{M}^{(a)} := \mathbf{a}^{(a)} \otimes \mathbf{a}^{(a)}$, $a = 1,2$ are the structural tensors characterizing the fiber directions. There exist different possibilities to model the mechanical response of soft biological tissue; see, e.g., [2, 12]. We are interested in polyconvex energy functions. For the construction of anisotropic, polyconvex functions, see, e.g., [18]. Here, we use the model due to [12], which was denoted model ψ_B in [3],

$$\begin{aligned}\psi = {} & c_1\left(I_1 I_3^{-1/3} - 3\right) + \sum_{a=1}^{2} \frac{k_1}{2k_2}\left\{\exp\left(k_2\left\langle J_4^{(a)} I_3^{-1/3} - 1\right\rangle^2\right) - 1\right\} \\ & + \varepsilon_1\left(I_3^{\varepsilon_2} + I_3^{-\varepsilon_2} - 2\right)^{\alpha},\end{aligned}$$

with the invariants $I_1 = \mathrm{tr}\,\mathbf{C}, I_2 = \mathrm{tr}[\mathrm{Cof}(\mathbf{C})], I_3 = \det\mathbf{C}, J_4^{(a)} = \mathrm{tr}[\mathbf{C}\mathbf{M}^{(a)}]$, $J_5^{(a)} = \mathrm{tr}[\mathbf{C}^2\mathbf{M}^{(a)}]$. Here, $\langle\bullet\rangle$ denote the Macauly brackets, $\langle\bullet\rangle = (|\bullet| + \bullet)/2$. The penalty term $\varepsilon_1\left(I_3^{\varepsilon_2} + I_3^{-\varepsilon_2} - 2\right)^{\alpha}$ models the incompressibility.

We adjust our parameters to experimental results in [11]; for details, see [5]. The adjustment results in the parameters $c_1 = 7.17\,[\mathrm{kPa}], k_1 = 3.69e-3\,[\mathrm{kPa}], k_2 = 51.2$ for the adventitia and $c_1 = 9.23\,[\mathrm{kPa}], k_1 = 193\,[\mathrm{kPa}], k_2 = 2.627e3$ for the media.

In the augmented Lagrange approach [10, 20] a Lagrange multiplier is introduced on each finite element and $\mu^T(\det\mathbf{F} - \mathbf{1})$ is added to the energy ψ. Here, we mean by $\det\mathbf{F}$ the vector of element-wise determinants of $\mathbf{F}$. The Lagrange multiplier will be computed iteratively by an Uzawa-like iteration $\mu_{k+1} = \mu_k + \xi_k(\det\mathbf{F} - \mathbf{1})$, where in our computations in Sect. 3 the series ξ_k will be chosen as a constant $\xi_k = \xi = 499.0$. We have chosen ξ by hand from the set $\{99, 499, 999, 1999, 9999\}$.

Our parameter fit is performed assuming incompressibility of the material. When using the penalty approach we have to choose sufficiently large penalty parameters. Here, our penalty parameters are $\varepsilon_1 = 70.0\,[\mathrm{kPa}], \varepsilon_2 = 8.5, \alpha = 1$ for the adventitia and $\varepsilon_1 = 360.0\,[\mathrm{kPa}], \varepsilon_2 = 9.0, \alpha = 1$ for the media. Also in the augmented Lagrange approach we need to choose our penalty parameters but here the penalty may be relaxed significantly, i.e., we choose $\varepsilon_1 = 10.0\,[\mathrm{kPa}], \varepsilon_2 = 4.0, \alpha = 1$ for adventitia and media. The relaxation becomes evident when the penalty function is plotted for the different sets of parameters. A sufficiently accurate stopping criterion has to be chosen for the augmented Lagrange loop; here we chose a tolerance of $|\det(\mathbf{F}) - 1| \leq 0.01$ on each element.

In our discretization, we have to avoid locking effects. We therefore replace point-wise penalization by the penalization of the average volumetric change on every finite element. This is accomplished, as in [3, 16], by applying a three-field formulation, known as the $\bar{\mathbf{F}}$-approach; see [19]. We use 10-noded tetrahedral elements for the displacement.

In our nonlinear scheme we solve a sequence of linear problems obtained from Newton's method, see, e.g., Fig. 2. This is also referred to as (pseudo) time stepping or load stepping. To obtain a fair comparison, we have chosen an automatic time stepping strategy. For the penalty approach we increase Δt when the number of Newton iterations is smaller than 6 and decrease Δt when it is larger than 9. This choice produced the best results. The simultaneous Augmented Lagrange approach, where the iteration for the Lagrange multiplier simultaneously to the Newton correction, can be viewed as an inexact Newton method see Fig. 3. Thus, a quadratic convergence cannot be expected. We therefore have chosen the bounds for the auto time stepping as 18 and 36. For all approaches the maximal time step size was bounded by $\Delta t_{\max} = 0.4$.

Fig. 2. Penalty for the incompressibility

Nonlinear Iteration (Penalty)

Set $k = 0$ and $t_0 = \Delta t_0$;
Apply partial load $t_k \cdot \mathbf{f}_{\text{load}}$ if the full load is not yet reached;
  Use Newton iteration to solve the nonlinear problem.
    Use GMRES to solve linearized problem using the
    FETI domain decomposition method as a preconditioner;
    Apply Newton correction;
  Adapt load step size Δt_{k+1}, i.e.,
  $\Delta t_{k+1} = 10^{1/5}\Delta t_k$, $\Delta t_{k+1} = 10^{-1/5}\Delta t_k$, or $\Delta t_{k+1} = \Delta t_k$;
  Set $t_{k+1} = t_k + \Delta t_{k+1}$;

2 FETI-DP Method

We briefly introduce the well-known FETI-DP method. For a more detailed introduction, see, e.g., [13, 16, 17, 21]. For algorithms of the **F**inite **E**lement **T**earing and **I**nterconnecting-type (FETI); see [6–9]. Using FETI-DP methods linear systems with billions of unknowns have been solved, e.g., in [14, 16] on large parallel machines.

We decompose the domain Ω into N nonoverlapping subdomains Ω_i. For all subdomains Ω_i, we assemble the local stiffness matrices $\mathbf{K}^{(i)}$ and local load vectors $\mathbf{f}^{(i)}$, $i = 1, \ldots, N$,

Fig. 3. Simultaneous augmented Lagrange for the incompressibility [10, 20]

Nonlinear Iteration (Simultaneous Augmented Lagrange)

Set $k = 0$ and $t_0 = \Delta t_0$;
Apply partial load $t_k \cdot \mathbf{f}_{\text{load}}$ if the full load is not yet reached;
  Set Lagrange parameter $\mu_0 = 0$;
  While Newton iteration has not converged and while elements with
  $|\det(\mathbf{F}) - 1| \geq$ TOL exist: Solve nonlinear problem with simultaneous
  Newton iteration and iteration for μ
    Use GMRES to solve linearized problem using the FETI method
    Apply Newton correction and update Lagrange parameter $\mu_{k+1} = \mu_k + \xi_k(\det\mathbf{F} - 1)$;
  Adapt load step size Δt_{k+1}, i.e.,
  $\Delta t_{k+1} = 10^{1/5}\Delta t_k$, $\Delta t_{k+1} = 10^{-1/5}\Delta t_k$, or $\Delta t_{k+1} = \Delta t_k$.
  Set $t_{k+1} = t_k + \Delta t_{k+1}$;

$$\mathbf{K} = \begin{bmatrix} \mathbf{K}^{(1)} & & \\ & \ddots & \\ & & \mathbf{K}^{(N)} \end{bmatrix}, \mathbf{u} = \begin{bmatrix} \mathbf{u}^{(1)} \\ \vdots \\ \mathbf{u}^{(N)} \end{bmatrix}, \mathbf{f} = \begin{bmatrix} \mathbf{f}^{(1)} \\ \vdots \\ \mathbf{f}^{(N)} \end{bmatrix}.$$

The interface is $\Gamma = \cup_{i=1}^{N} \partial\Omega_i \setminus \partial\Omega$. The discrete problem can be formulated as minimization problem with the interface continuity constraint $\mathbf{Bu} = \mathbf{0}$, where $\mathbf{B} = [\mathbf{B}^{(1)}, \ldots, \mathbf{B}^{(N)}]$ with entries from $0, 1, -1$. By introducing Lagrange multipliers λ to enforce the continuity along the subdomain interface we obtain the problem: Find $(\mathbf{u}, \lambda)$, such that

$$\begin{aligned} \mathbf{Ku} + \mathbf{B}^T\lambda &= \mathbf{f} \\ \mathbf{Bu} &= \mathbf{0}. \end{aligned}$$

This problem can be solved by eliminating the displacement variables $\mathbf{u}$ and solving the resulting Schur complement system by conjugate gradients.

In FETI-DP methods some continuity constraints are enforced on *primal* displacement variables $\tilde{\mathbf{u}}_\Pi$ throughout iterations to enforce invertibility of the local problems. This yields a saddle point problem of the form

$$\begin{aligned} \widetilde{\mathbf{K}}\tilde{\mathbf{u}} + \mathbf{B}^T\lambda &= \tilde{\mathbf{f}} \\ \mathbf{B}\tilde{\mathbf{u}} &= \mathbf{0}, \end{aligned}$$

where the matrix $\widetilde{\mathbf{K}}$ and right hand side $\tilde{\mathbf{f}}$ are partially assembled in the primal variables, i.e.,

$$\widetilde{\mathbf{K}} = \begin{bmatrix} \mathbf{K}_{BB}^{(1)} & & \widetilde{\mathbf{K}}_{\Pi B}^{(1)T} \\ & \ddots & \vdots \\ & \mathbf{K}_{BB}^{(N)} & \widetilde{\mathbf{K}}_{\Pi B}^{(N)T} \\ \widetilde{\mathbf{K}}_{\Pi B}^{(1)} \cdots & \widetilde{\mathbf{K}}_{\Pi B}^{(N)} & \widetilde{\mathbf{K}}_{\Pi\Pi} \end{bmatrix}, \qquad \tilde{\mathbf{f}} = \begin{bmatrix} \mathbf{f}_B^{(1)} \\ \vdots \\ \mathbf{f}_B^{(N)} \\ \tilde{\mathbf{f}}_\Pi \end{bmatrix}.$$

The coupling also provides the coarse problem for the method. Reducing the system of equations to an equation in λ, it remains to solve iteratively

$$\mathbf{M}_{\mathrm{D}}^{-1}\mathbf{F}_{feti}\lambda = \mathbf{M}_{\mathrm{D}}^{-1}\mathbf{d},$$

where $\mathbf{F}_{feti} = \mathbf{B}\widetilde{\mathbf{K}}^{-1}\mathbf{B}^T$, and $\mathbf{M}_{\mathrm{D}}^{-1} = \mathbf{B}_D\mathbf{R}_\Gamma^T\mathbf{S}\mathbf{R}_\Gamma\mathbf{B}_D^T$ is the Dirichlet preconditioner. Here, $\mathbf{S}$ is the Schur complement obtained by eliminating the interior variables in every subdomain, i.e., $\mathbf{S} = \begin{bmatrix} \mathbf{S}^{(1)} & & \\ & \ddots & \\ & & \mathbf{S}^{(N)} \end{bmatrix}$. The operator $\mathbf{R}_\Gamma$ is a restriction matrix, consisting of zeros and ones, that, when applied to a vector $\tilde{\mathbf{u}}$, removes the interior variables from $\tilde{\mathbf{u}}$. The matrices $\mathbf{B}_D$ are scaled variants of the jump operator $\mathbf{B}$ where, in the simplest case, the contribution from and to each interface node is scaled by the inverse of the multiplicity of the node. We define the multiplicity of a node as the number of subdomains it belongs to. For heterogeneous problems a more elaborate scaling, using an appropriate scaling factor, defined by the coefficients ρ_i, is necessary; see, e.g., [17, p. 1532, Formula (4.3)] and [15, p. 1403, Formula (6)].

3 Numerical Results

A pressure of 200 mmHg is applied to the inside of the artery, see Fig. 1. The FETI-DP iteration is stopped when the absolute residual is reduced to 5×10^{-9}; we have 224 subdomains. The total cost can be estimated by multiplying the number of Newton steps by the corresponding average number of (inner) FETI-DP Krylov iterations, see Tables 1 and 2.

Our results show that the use of the augmented Lagrange method can significantly improve the properties of the linearized systems occurring in the nonlinear solution scheme. The convergence of the nonlinear scheme is also improved, i.e., in our nonlinear scheme larger pseudo time steps Δt can be chosen. Of course, an additional iteration process for the Lagrange multiplier is introduced. Here, this iteration process is carried out simultaneously with the Newton iteration.

The results in Tables 1 and 2 show that the additional cost for the augmented Lagrange iteration is more than amortized by the faster convergence of the nonlinear scheme and the linear iterative solver. Moreover, in the augmented Lagrange approach the volumetric change is exactly controlled during the iteration process, i.e., we have satisfied element-wise the condition $|\det(\mathbf{F}) - 1| \leq 0.01$. In the penalty approach the volumetric change produced by the chosen penalty parameters is only

Table 1. Newton iteration for the penalty formulation. Pseudo-time t, number of Newton steps, average number of Krylov iterations per Newton step.

t	Newton steps	∅ Krylov its
0.010	9	172.2
0.020	5	173.0
0.036	5	175.8
0.061	5	179.4
0.101	6	189.3
0.141	5	187.0
0.204	6	201.8
0.267	5	195.6
0.367	7	208.0
0.467	7	204.1
0.567	5	207.4
0.725	6	217.8
0.884	5	225.4
1.135	6	242.0
1.386	6	253.8
1.637	7	266.3
1.889	5	279.4
2.000	4	285.8
	Σ 104	Total ∅ 213.3

Table 2. Simultaneous Newton and augmented Lagrange (AL) iteration. Pseudo-time t, number of Newton-AL steps, average number of Krylov iterations per Newton-AL step.

t	Newton-AL steps	∅ Krylov its
0.010	9	99.3
0.026	4	100.5
0.051	5	101.4
0.091	6	101.3
0.154	6	102.8
0.254	7	104.3
0.412	11	105.4
0.664	14	109.4
1.062	14	119.0
1.462	16	139.7
1.862	17	167.0
2.000	15	180.8
	Σ 124	Total ∅ 138.6

known ex-post. In our example the solution using the penalty approach only satisfies $|\det(\mathbf{F}) - 1| \leq 0.021$.

In the results in Table 2, we see that the number of Newton-AL-iterations increases during the simulation. This is due to the fact that in the beginning of the simulation only a very small number of finite elements violate the element-wise condition $|\det(\mathbf{F}) - 1| \leq 0.01$.

The results in, both, Tables 1 and 2 also show an increase of the FETI-DP iterations during the simulation. We believe that this may in part be due to an increasing influence of the incompressibility constraint during the simulation but also result from the exponential stiffening behavior of the fibers. In [1], we have observed that the anisotropies introduced to the material wall models by the terms modeling the fibers can have a visible impact on the convergence of the nonlinear iteration scheme as well as the convergence of the iterative linear solver. Ideas described in [16] may improve the convergence of domain decomposition solvers for such anisotropic problems.

Bibliography

[1] Daniel Balzani, Dominik Brands, Axel Klawonn, Oliver Rheinbach, and Jörg Schröder. On the mechanical modeling of anisotropic biological soft tissue and iterative parallel solution strategies. *Arch. Appl. Mech.*, 80(5):479–488, 2010.

[2] Daniel Balzani, Patrizio Neff, Jörg Schröder, and Gerhard A. Holzapfel. A polyconvex framework for soft biological tissues. Adjustment to experimental data. *Int. J. Solids Struct.*, 43(20):6052–6070, 2006.

[3] Dominik Brands, Axel Klawonn, Oliver Rheinbach, and Jörg Schröder. Modelling and convergence in arterial wall simulations using a parallel FETI solution strategy. *Comput. Methods Biomec.*, 11:569–583, 2008.

[4] Dominik Brands, Jörg Schröder, Axel Klawonn, Oliver Rheinbach, Dirk Böse, and Raimund Erbel. Numerical simulations of arterial walls based on ivus-data. *PAMM*, 9(1):75–78, 2009.

[5] Sarah Brinkhues, Axel Klawonn, Oliver Rheinbach, and Jörg Schröder. Parallel simulation of arterial walls. 2011. In preparation.

[6] Charbel Farhat, Michel Lesoinne, Patrick LeTallec, Kendall Pierson, and Daniel Rixen. FETI-DP: A dual-primal unified FETI method - part i: A faster alternative to the two-level FETI method. *Internat. J. Numer. Methods Engrg.*, 50:1523–1544, 2001.

[7] Charbel Farhat, Michel Lesoinne, and Kendall Pierson. A scalable dual-primal domain decomposition method. *Numer. Lin. Alg. Appl.*, 7:687–714, 2000.

[8] Charbel Farhat, Jan Mandel, and Francois-Xavier Roux. Optimal convergence properties of the FETI domain decomposition method. *Comput. Methods Appl. Mech. Engrg.*, 115:367–388, 1994.

[9] Charbel Farhat and Francois-Xavier Roux. A method of Finite Element Tearing and Interconnecting and its parallel solution algorithm. *Int. J. Numer. Meth. Engrg.*, 32:1205–1227, 1991.

[10] M. Fortin and A. Fortin. A generalization of Uzawa's algorithm for the solution of the Navier-Stokes equations. *Comm. Appl. Numer. Methods*, 1(5):205–208, 1985.

[11] Gerhard A. Holzapfel. Determination of material models for arterial walls from uniaxial extension tests and histological structure. *J. Theor. Biol.*, 238(2):290–302, 2006.

[12] Gerhard A. Holzapfel, Thomas C. Gasser, and Ray W. Ogden. A new constitutive framework for arterial wall mechanics and a comparative study of material models. *J. Elasticity*, 61(1–3):1–48, 2000.

[13] Axel Klawonn and Oliver Rheinbach. A parallel implementation of Dual-Primal FETI methods for three dimensional linear elasticity using a transformation of basis. *SIAM J. Sci. Comput.*, 28:1886–1906, 2006.

[14] Axel Klawonn and Oliver Rheinbach. Inexact FETI-DP methods. *Inter. J. Numer. Methods Engrg.*, 69:284–307, 2007.

[15] Axel Klawonn and Oliver Rheinbach. Robust FETI-DP methods for heterogeneous three dimensional linear elasticity problems. *Comput. Methods Appl. Mech. Engrg.*, 196:1400–1414, 2007.

[16] Axel Klawonn and Oliver Rheinbach. Highly scalable parallel domain decomposition methods with an application to biomechanics. *ZAMM - Z. Angew. Math. Mech.*, 90(1):5–32, July 2010.

[17] Axel Klawonn and Olof B. Widlund. Dual-Primal FETI Methods for Linear Elasticity. *Comm. Pure Appl. Math.*, 59:1523–1572, 2006.

[18] Jörg Schröder and Patrizio Neff. Invariant formulation of hyperelastic transverse isotropy based on polyconvex free energy functions. *Int. J. Solids Struct.*, 40:401–445, 2003.

[19] Juan C. Simo. Numerical analysis and simulation of plasticity. In P.G. Ciarlet and J.L. Lions, editors, *Handbook of numerical analysis*, volume IV. Elsevier, 1998.

[20] Juan C. Simo and Robert L. Taylor. Quasi-incompressible finite elasticity in principal stretches. continuum basis and numerical algorithms. *Comput. Methods Appl. Mech. Engrg.*, 85(3):273–310, 1991.

[21] Andrea Toselli and Olof B. Widlund. *Domain Decomposition Methods - Algorithms and Theory*, volume 34 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin Heidelberg New York, 2005.

Techniques for Locally Adaptive Time Stepping Developed over the Last Two Decades

Martin J. Gander[1], Laurence Halpern[2]

[1] University of Geneva, CH-1211, Geneva, Switzerland martin.gander@unige.ch
[2] Université Paris 13, CNRS, UMR 7539 LAGA, 99 av. Jean-Baptiste Clément, 93430 7 Villetaneuse, France halpern@math.univ-paris13.fr

Adaptive mesh refinement techniques are well established and widely used for space discretizations. In contrast, local time stepping is much less used, and the corresponding techniques are less mature, needing delicate synchronization steps, which involve interpolation, extrapolation or projection. These operations can have adverse effects on the stability, and can also destroy important geometric properties of the scheme, like for example the conservation of invariants. We give here a survey on the intensive research performed in this direction over the last two decades.

1 Methods from the ODE Community

Local time stepping started in the ODE community with the development of split Runge-Kutta methods with Rice [34]. Nowadays called multirate Runge-Kutta methods, these methods were first developed for naturally split systems of ordinary differential equations $y' = b(y,z,t)$ and $z' = c(y,z,t)$, in which the z components need to be integrated on a finer time mesh than the y components. One then uses a Runge-Kutta method for the fast, so called active components with a small time step, and another one for the slow, so called latent components, with a large time step, and uses either interpolation or extrapolation for the missing values, depending on which of the components are computed first, see [27].

Multirate time integration methods were also proposed for linear multistep methods in [22], with two main approaches: fastest-first and slowest-first. Suppose an implicit linear multistep method is used. In the fastest-first approach, one advances the z components with small time steps h, and whenever one needs a component of the slow part y, one uses a predictor step for it. Once the fine stepping scheme arrives at a coarse step H, the slow solution component y is also computed. The major disadvantage of this approach is that it is very difficult to do adaptive time stepping. This is easier in the slowest-first approach, where first the slow component is doing an adaptive integration step, until one is accepted with step size H. Then the adaptive fine integration is tried with small steps h, until one reaches with several accepted small steps the coarse level H. For the slow adaptive step H however, one needs also

R. Bank et al. (eds.), *Domain Decomposition Methods in Science and Engineering XX*, Lecture Notes in Computational Science and Engineering 91,
DOI 10.1007/978-3-642-35275-1_44,

an approximation of the fast component for coupled components, and the authors in [22] say: "There are several possible ways to control the fast extrapolation error, none of which is entirely satisfactory". The stability properties of such multirate schemes were analyzed in [35] for Backward Euler multirate schemes; see also [23].

In contrast to the multirate methods, multirate extrapolation methods aim at integrating systems of ODEs without a priori knowledge of which components need finer time integration steps than others. A method based on Richardson extrapolation was proposed in [13]: one computes approximations for all components for a time step sequence $\{h_1, h_2, h_3, \ldots\}$, e.g. $h_2 = \frac{h_1}{2}$, $h_3 = \frac{h_1}{3}, \ldots$, and then builds the Richardson extrapolation table. As soon as a component has reached the desired accuracy at step h_k (an error estimate is available automatically in the Richardson table), extrapolation for this component is marked inactive, and only components needing further accuracy continue the extrapolation. Inactive components must then however be approximated in order for the extrapolation to continue. Using interpolation from the continuous approximation obtained from the Richardson extrapolation can completely destroy the extrapolation process, which is based on the same error expansion for all the components. The authors in [13] propose instead an elegant approximation from the asymptotic expansion assumption itself, and also introduce a defect control to avoid that inactivation fails in certain situations.

2 Methods from the PDE Community

Local time stepping schemes in the PDE community started with experimental work, see for example [28]. Such ad hoc solutions were quite different for parabolic and hyperbolic PDEs.

Hyperbolic Problems: a first complete mathematical analysis of two space-time adaptive schemes for the wave equation $u_t = u_x$, an interpolation based variant, and the so called coarse mesh approximation method were given by Berger [2] (see also [3], and an early analysis for a different technique based on finite volumes in [31]). Using for example a three point explicit scheme, the interpolation based approach starts with a coarse step at the interface, shown in red in Fig. 1 on the left, followed by an interpolation for the fine grid values, shown in blue. In the coarse mesh approximation, one uses the coarse spatial mesh to compute small time steps Δt, $2\Delta t$, $3\Delta t, \ldots$ at the interface, instead of interpolating these values, as indicated in Fig. 1 on the right for the second step $2\Delta t$ in red, where the blue value at Δt has already been

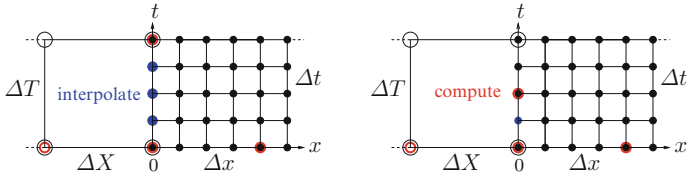


Fig. 1. Interpolation based approach on the *left*, and the coarse mesh method on the *right*

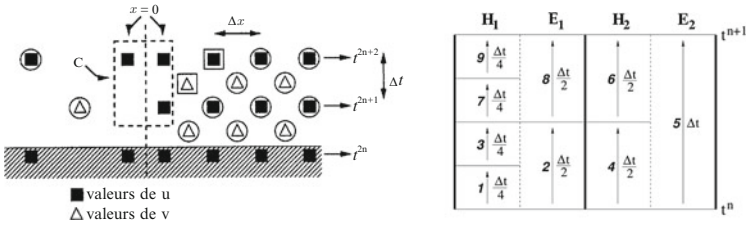


Fig. 2. First energy-preserving local time stepping for the wave equation on the *left*, and symplectic scheme for Maxwell's equation on the *right*

computed. The author proves for the hyperbolic model problem $u_t = u_x$ that both approaches are stable for the Lax-Wendroff scheme, but stability for the Leapfrog scheme can only be achieved with overlap. Elegant recursive versions of such algorithms are in [33].

A key new ingredient to obtain stability for a Leapfrog type scheme for the locally adaptive solution of the wave equation can be found in the seminal papers by Collino et al. in [7, 8]: the introduction of a discrete energy conservation. In presentations, this approach was always introduced with an impressive movie, where a wave passes a locally refined patch, and everything looks fine for quite a long time after the wave has passed, until suddenly an instability forms at the boundary of the patch, and the numerical solution explodes, if a simple interpolation based scheme is used. The method was first described for the 1d Maxwell system $u_t + v_x = 0$, $v_t + u_x = 0$, which is equivalent to the 1d second order wave equation $u_{tt} = u_{xx}$, and can best be described with the original picture from [7] shown in Fig. 2 on the left. Thinking just about the second order wave equation, discretized with a centered finite difference scheme both in space and time, we get the five point star, well visible with the black squares in Fig. 2 (the triangles would be for the unknowns v we do not consider here). Now all points can be computed with this star at time levels t^{2n+1} and t^{2n+2}, given the values at earlier time levels, except for the values in the dashed box. The key idea of the energy preserving scheme is now to permit two different values at $x = 0$ at even time levels t^{2n}, and to introduce as additional equation the discrete energy, which needs to be preserved. This leads naturally to a stable scheme, but it requires the solution of a small linear system at the interface. Energy conservation turned out to be a key tool for stability analysis, and is used now for other space-time adaptive methods, see for example [11], where the authors introduce an unusual energy, in order to analyze the stability of their space-time locally adaptive scheme.

A very elegant way of generalizing a symplectic integrator (which naturally preserves a nearby energy) for variable step size integration was presented in [26], and adapted to Maxwell's system in [32]. The Störmer-Verlet scheme is symplectic for these equations, and is shown in Fig. 2 on the right. Without refinement, the scheme is visible in the right part under H_2, E_2: we see that first a half step denoted by 4 is performed for the magnetic field H, followed by a full step denoted by 5 for the electric field E, and concluded by a second half step for H denoted by 6. In each of these steps, the Störmer-Verlet scheme uses for H the newest values available from

the other field E, and vice versa. It turns out that doing the same over the locally refined region shown in Fig. 2 on the right, and performing the steps in the given order, starting with 1 and ending with 9, and using each time the newest information available, is still symplectic! Since symplectic schemes preserve a nearby energy, this scheme has all the good stability properties needed.

In a finite volume or discontinuous Galerkin in the time domain setting (DGTD), on unstructured meshes in space, the scheme in each subdomain with given time step can be advanced until the new time value reaches that of its neighbor, according to the stability constraint, see [12] for elastodynamics computations in the context of ADER methods (Arbitrary high order, using high order DERivatives of polynomials).

Parabolic Problems are often integrated using implicit methods, which require the solution of large systems of equations. These systems are obtained using the same time step over the entire domain, and it is thus a priori not possible to use a local time step. The first ideas to change this are based on domain decomposition methods, where then interface values have to be predicted in some way, before the subdomain problems are advanced in time by an implicit method.

A first interesting way to explicitly predict the interface values appeared in [9], where a third spatial discretization size H is introduced, in addition to h_l and h_r, see Fig. 3 on the left. The method then first does an explicit prediction step over the big Δt, stable because the corresponding spatial step H is big, as indicated in red. This is followed by interpolation (in blue) to obtain all needed values at the interface, and then on each side one can do implicit solves to advance the method. It is proved in [9] that this scheme is stable for the heat equation with a centered finite difference discretization in space, and forward/backward Euler in time, if $\Delta t \leq \frac{1}{2}H^2$, and the error satisfies the estimate $\max|\text{err}| \leq C(h_l^2 + h_r^2 + H^3 + \Delta t_l + \Delta t_r + H\Delta t)$, which shows impressively that the big prediction step Δt, H only affects the accuracy in higher order terms!

A different approach was proposed by Blum et al. [4], as shown in Fig. 3 on the right. The authors do not consider local refinement in time and space, their main interest is to break up a large linear system from the implicit time integration into smaller ones, but their idea can also be used for local adaptation in time and space. The key idea is to use overlap, predict all values needed at the interfaces using a higher order extrapolation method, and then solve implicitly on the corresponding subdomains to advance the method. The authors prove for the heat equation without

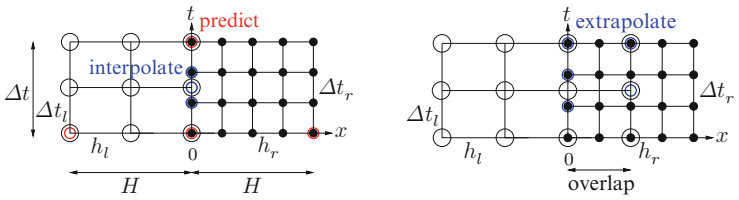


Fig. 3. Explicit prediction of the interface values on an intermediate spatial grid on the *left*, and by extrapolation with overlap on the *right*

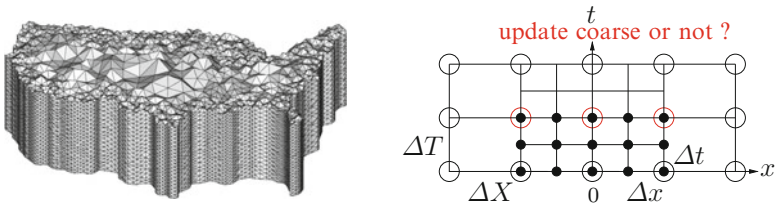


Fig. 4. A completely general space time mesh on the *left*, and the one-way and two way approaches on the *right*

local refinement, $h_l = h_r = h$ and $\Delta t_l = \Delta t_r = \Delta t$, that the Crank-Nicolson scheme is stable, provided that $\Delta t \leq C \left(\frac{L}{\log L}\right)^2 h^2$, where Lh is the overlap, and an error estimate of the form $O(\Delta t^2 + h^2)$. So here increasing the overlap can lessen the stability constraint on the time step.

If one wants to avoid any time step constraints, one can perform the coupling fully implicitly, as proposed in [16]. Here, one simply writes the implicit scheme on the fine and coarse subdomain, and the interpolation conditions into one big system of linear equations, which is then solved. The authors show for a linear advection reaction diffusion equation that a standard centered scheme with backward Euler in time is unconditionally stable, and satisfies for $\Delta t = O(h)$ the error estimate $O(\Delta t + h^2)$ in 1d, but in 2d there is a loss of $|\log h|^{\frac{1}{2}}$, and in 3d a loss of $\frac{1}{\sqrt{h}}$ in accuracy.

A more general approach based on domain decomposition can be found in [17]. For the heat equation $u_t = u_{xx}$, and the decomposition of the domain $\Omega = (-1,1)$ into two subdomains $\Omega_1 = (-1,0)$ and $\Omega_2 = (0,1)$, the authors propose to discretize the coupling conditions $u_1(0) = u_2(0)$, $\partial_x u_1(0) = \partial_x u_2(0)$ using a conservative finite volume discretization over non-matching time grids. They also obtain, for each variant of the method, a very large system of equations to solve, but propose to solve it using one or several steps of an iterative Dirichlet-Neumann algorithm. They show that these schemes are conservative, provided one stops the iteration after a Neumann step, and satisfy an error estimate $O(\Delta t + h)$ under certain conditions. One can show that one of their methods corresponds to the approach in [16].

Space-Time Finite Element Methods consider the time direction like one of the spatial directions, and discretize the problem directly in space-time by a finite element method, which leads to a large discrete problem in space-time. These methods have their roots in the work of C. Johnson and co-authors, see for instance [15] for a review. Discontinuous Galerkin methods were used, and the adaptation was done through a posteriori estimates. In the first versions of the method, the space-time finite elements were still special, since they always had boundaries in time aligned with the time direction, for example prisms. Completely general triangular meshes in space time require special meshing techniques, since they need to satisfy certain angle constraints, in order to avoid total global coupling in space-time, see [36] for applications to Burger's equation and elastodynamics. An impressive example of such a mesh from [14] is reproduced in Fig. 4 on the left. A very recent contribution

using discontinuous Galerkin methods can be found in these proceedings, see [30].

One-Way and Two-Way Methods are in principle very different from all the methods we considered earlier, since they have both a coarse *and* a fine mesh in parts of the domain. They have their roots in weather and climate simulations, which often use a global model over a large region, for example the entire planet, and then refined models over a small region, for example a country. The question is then how to compute a refined solution based on the solution of the global coarse problem. In [10] and [6], the so called one-way (or "offline") and two-way (or "online") methods are proposed. In the one-way method, the coarse model is first solved once and for all, and stored. Then boundary data is extracted to be imposed on the boundary of the smaller refined region. The simplest approach is to use Dirichlet conditions, which can however lead to large errors. A more refined approach is to use so called open boundary conditions, which are related to absorbing boundary conditions, but different, see [6, 29]. Open boundary conditions lead in general to substantially more accurate fine models. In the two way approach, one only performs one or a few time steps of the coarse model, then solves the fine model in the refined region as before, but updates the coarse result whenever a more accurate fine result is available, before continuing the next coarse time step, see Fig. 4 on the right. If one simulates only one time step of the coarse model before solving the fine model and uses Dirichlet conditions, this approach is very much related to the first approach for hyperbolic problems described earlier.

Schwarz waveform relaxation methods are the most flexible methods for solving evolution problems locally adaptively in space time, since they permit not only refined time steps, but even different numerical methods, or different models in different regions. They were first described in [20] and are based on a decomposition in space of the domain over which the evolution problem is posed and a subdomain iteration in space-time: starting with an initial guess on each space-time interface between subdomains, on each subdomain the evolution problem is solved over an entire so called time window. Then information is exchanged between subdomains using transmission conditions, and the subdomain problems are solved again and again until a suitable matching is reached. So the price to pay for this flexibility and generality is the iteration. The method from [17] we have seen earlier is in this class of methods, but much faster convergence can be obtained when optimized transmission conditions are used, see [1, 18, 21, 24, 25], and references therein. Very general non-matching space-time grids can be coupled like this using a projection algorithm with optimal linear complexity from [19]. For recent realistic applications in a complex setting, see [5].

Bibliography

[1] D. Bennequin, M. J. Gander, and L. Halpern. A homographic best approximation problem with application to optimized Schwarz waveform relaxation. *Math. of Comp.*, 78(265):185–232, 2009.

[2] M. Berger. Stability of interfaces with mesh refinement. *Math. of Comp.*, 45: 301–318, 1985.
[3] M. J. Berger and J. Oliger. Adaptive mesh refinement for hyperbolic partial differential equations. *J. Comp. Phys.*, 53:484–512, 1984.
[4] H. Blum, S. Lisky, and R. Rannacher. A domain splitting algorithm for parabolic problems. *Computing*, 49:11–23, 1992.
[5] M. Borrel, L. Halpern, and J. Ryan. Euler - Navier-Stokes coupling for aeroacoustics problems. In A. Kuzmin, editor, *Computational Fluid Dynamics 2010, ICCFD6*, pages 427–434. Springer-Verlag, 2010.
[6] S. Cailleau, V. Fedorenko, B. Barnier, E. Blayo, and L. Debreu. Comparison of different numerical methods used to handle the open boundary of a regional ocean circulation model of the bay of biscay. *Ocean Modelling*, 25:1–16, 2008.
[7] F. Collino, T. Fouquet, and P. Joly. A conservative space-time mesh refinement method for the 1d wave equation. part I: construction. *Numer. Math.*, 95:197–221, 2003.
[8] F. Collino, T. Fouquet, and P. Joly. A conservative space-time mesh refinement method for the 1d wave equation. part II: analysis. *Numer. Math.*, 95:223–251, 2003.
[9] C. Dawson, Q. Du, and T. Dupont. A finite difference domain decomposition algorithm for numerical solution of the heat equation. *Math. Comp.*, 57(195): 63–71, 1991.
[10] L. Debreu and E. Blayo. Two-way embedding algorithms: a review. *Ocean Dynamics*, 58:415–428, 2008.
[11] J. Diaz and M. J. Grote. Energy conserving explicit local time-stepping for second-order wave equations. *SIAM J. Scientific Computing*, 31:1985–2014, 2009.
[12] M. Dumbser, M. Käser, and E.F. Toro. An arbitrary high-order discontinuous Galerkin method for elastic waves on unstructured meshes V: Local time stepping and p-adaptivity. *Geophysical Journal International*, 171(2):695–717, 2007.
[13] Ch. Engstler and Ch. Lubich. Multirate extrapolation methods for differential equations with different time scales. *Computing*, 58:173–185, 1997.
[14] J. Erickson, D. Guoy, J.M. Sullivan, and A. Üngör. Building spacetime meshes over arbitrary spatial domains. *Eng. with Comp.*, 20:342–353, 2005.
[15] K. Eriksson, C. Johnson, and A. Logg. Adaptive computational methods for parabolic problems. In *Encyclopedia of Computational Mechanics*, 2004.
[16] R. E. Ewing, R. D. Lazarov, and A. T. Vassilev. Finite difference scheme for parabolic problems on composite grids with refinement in time and space. *SIAM J. Numer. Anal.*, 31:1605–1622, 1994.
[17] I. Faille, F. Nataf, F. Willien, and S. Wolf. Two local time stepping schemes for parabolic problems. *ESAIM: proceedings*, 29:58–72, 2009.
[18] M. J. Gander and L. Halpern. Optimized Schwarz waveform relaxation methods for advection reaction diffusion problems. *SIAM J. Numer. Anal.*, 45(2): 666–697, 2007.

[19] M. J. Gander and C. Japhet. An algorithm for non-matching grid projections with linear complexity. In M. Bercovier, M.J. Gander, D. Keyes, and O.B. Widlund, editors, *Domain Decomposition Methods in Science and Engineering XVIII*, pages 185–192. Springer Verlag LNCSE, 2008.

[20] M. J. Gander and A. M. Stuart. Space-time continuous analysis of waveform relaxation for the heat equation. *SIAM J. Sci. Comput.*, 19(6):2014–2031, 1998.

[21] M. J. Gander, L. Halpern, and F. Nataf. Optimal Schwarz waveform relaxation for the one dimensional wave equation. *SIAM Journal of Numerical Analysis*, 41(5):1643–1681, 2003.

[22] C.W. Gear and D.R. Wells. Multirate linear multistep methods. *BIT*, 24:484–502, 1984.

[23] M. Günter, A. Kværnø, and P. Rentrop. Multirate partitioned Runge-Kutta methods. *BIT*, 38(2):101–112, 1998.

[24] L. Halpern. Non conforming space-time grids for the wave equation: a new approach. *Monografías del Seminario Matemático García de Galdeano*, 31: 479–495, 2004.

[25] L. Halpern. Local space-time refinement for the one dimensional wave equation. *J. of Comp. Acoustics*, 13(3):153–176, 2005.

[26] D.J. Hardy, D.I. Okunbor, and R.D. Skeel. Symplectic variable step size integration for N-body problems. *Appl. Numer. Math.*, 29(5):19–30, 1999.

[27] A. Kværno and P. Rentrop. Low order multirate Runge-Kutta methods in electric circuit simulation, 1999.

[28] R. Löhner, K. Morgan, and O. C. Zienkiewicz. The use of domain splitting with an explicit hyperbolic solver. *Computer Methods in Applied Mechanics and Engineering*, 45:313–329, 1984.

[29] V. Martin and E. Blayo. Revisiting the open boundary problem in computational fluid dynamics. In R. Bank, M. Holst, O.B. Widlund, and J. Xu, editors, *Domain Decomposition Methods in Science and Engineering XX*. Springer-Verlag, 2012.

[30] M. Neumüller and O. Steinbach. A DG Space-Time Domain Decomposition Method. In R. Bank, M. Holst, O.B. Widlund, and J. Xu, editors, *Domain Decomposition Methods in Science and Engineering XX*. Springer Verlag, 2013.

[31] S. Osher and R. Sanders. Numerical approximations to nonlinear conservation laws with locally varying time and space grids. *Math. of Comp.*, 41(164):321–336, 1983.

[32] S. Piperno. Symplectic local time-stepping in non-dissipative DGTD methods applied to wave propagation problems. *ESAIM: Mathematical Modelling and Numerical Analysis*, 40(5):815–841, 2006.

[33] F. Pretorius and L. Lehner. Adaptive mesh refinement for characteristic codes. *J. Comp. Phys.*, 198:10–34, 2004.

[34] R. C. Rice. Split Runge-Kutta methods for simulatneous equations. *J. Res. Natl. Bur. Standards*, 64B:151–170, 1960.

[35] S. Skelboe and P. U. Andersen. Stability properties of backward Euler multirate formulas. *SIAM J. Sci. Stat. Comp.*, 10:1000–1009, 1989.

[36] A. Üngör and A. Sheffer. Tent-pitcher: A meshing algorithm for space-time discontinuous galerkin methods. In *In proc. 9th int'l. meshing roundtable*, pages 111–122, 2000.

Newton-Schwarz Optimised Waveform Relaxation Krylov Accelerators for Nonlinear Reactive Transport

Florian Haeberlein[1,2], Laurence Halpern[2], and Anthony Michel[1]

[1] IFP Energies nouvelles, 1 et 4 avenue de Bois-Préau, 92852 Rueil-Malmaison, France, firstname.lastname@ifpen.fr

[2] Université Paris 13, CNRS, UMR 7539 LAGA, 99 av. Jean-Baptiste Clément, 93430 Villetaneuse, France, lastname@math.univ-paris13.fr

1 Introduction

Krylov-type methods are widely used in order to accelerate the convergence of Schwarz-type methods in the linear case. Authors in [2] have shown that they accelerate without overhead cost the convergence speed of Schwarz methods for different types of transmission conditions. In the nonlinear context, the well-known class of Newton-Krylov-Schwarz methods (cf. [5]) for steady-state problems or time-dependent problems uses the following strategy: time-dependent problems are discretised uniformly in time first and then one proceeds as for steady-state problems, i.e. the nonlinear problem is solved by a Newton method where the linear system at each iteration is solved by a Krylov-type method preconditioned by an algebraic Schwarz method. The major limitation is that NKS methods do not allow different time discretisations in the subdomains since the problem is discretised in time uniformly up from the beginning.

In this work, we are interested in applying the well-established technique from the linear case in the context of Schwarz Waveform Relaxation methods (SWR, cf. [8]) to nonlinear time-dependent problems in order to benefit from its accelerating properties. We emphasise the use of SWR methods since within this approach, it is possible to use different discretisations in time and space in the subdomains, even the coupling of different models is possible. In many applications, time step restrictions in implicit approaches are highly localised in space due to heterogeneity and SWR methods are perfectly suited to localise and isolate them in subdomains which are treated with different time discretisations.

Our motivation of balancing time step restrictions in the time-dependent nonlinear case on subdomains is close to the approach in [6, 11] where the balancing of nonlinearities on subdomains in the steady-state case is achieved using the permutation of domain decomposition methods and Newton's method in combination with Krylov accelerators.

The paper is organised as follows: In Sect. 2 we set up the problem to solve. In Sect. 3 we describe the Schwarz waveform relaxation (SWR) algorithm and the

R. Bank et al. (eds.), *Domain Decomposition Methods in Science and Engineering XX*, Lecture Notes in Computational Science and Engineering 91,
DOI 10.1007/978-3-642-35275-1_45,

reduction to the interface variables. The new approach is described in Sect. 4. Numerical issues and results are given in Sect. 5.

2 Problem Description

In this paper we consider the following model in $\Omega \times (0,T)$, $\Omega \subset \mathbb{R}^d$:

$$\partial_t(\phi w) + \mathscr{L}w + \mathscr{F}(w) = q \text{ in } \Omega \times (0,T), \tag{1}$$

$$w(\cdot,0) = w_0 \text{ in } \Omega, \quad \mathscr{G}w = g \text{ on } \partial\Omega \times (0,T). \tag{2}$$

where $\phi(x) > 0$ is the porosity, $w \in \mathbb{R}^s$ the vector containing the concentrations of the s chemical species. $\mathscr{L}[\cdot] = \nabla \cdot (-a\nabla + \mathbf{b})$ is a linear operator which models diffusion described by a positive scalar diffusion coefficient $a > 0$ and advection described by a Darcy field $\mathbf{b} \in \mathbb{R}^d$. The transport operator can be zero for non-mobile species. $\mathscr{F}$ is a nonlinear chemical coupling operator. We impose initial conditions on Ω given by w_0 and linear boundary conditions represented by $\mathscr{G}$, for instance Neumann or Dirichlet conditions. The data g and q are source terms depending on space and time.

3 The Schwarz Waveform Relaxation Algorithm and the Classical Approach

We decompose the domain Ω into two non-overlapping domains Ω_1 and Ω_2 and call the common boundary $\Gamma = \partial\Omega_1 \cap \partial\Omega_2$ the interface between the subdomains. We introduce the following SWR algorithm with Robin transmission conditions to approximate the solution of (1): given the iterate w_i^{k-1} which is equal to an initial guess for the first iteration, then one step of the algorithm consists in computing in parallel w_i^k for subdomains $\Omega_i = 1,2$, with data coming from the neighbouring subdomain $\Omega_\times$, with $\tilde{1} = 2$ and $\tilde{2} = 1$.

$$\partial_t(\phi w_i^k) + \mathscr{L}w_i^k + \mathscr{F}(w_i^k) = q \quad \text{in } \Omega_i \times (0,T), \tag{3}$$

$$(\partial_{n_i} + p)w_i^k = (\partial_{n_i} + p)w_\times^{k-1} \quad \text{on } \Gamma \times (0,T), \tag{4}$$

$$w_i^k(\cdot,0) = w_0 \text{ in } \Omega_i, \quad \mathscr{G}w_i^k = g \text{ on } \partial\Omega_i \setminus \Gamma \times (0,T), \tag{5}$$

with n_i the unit outward normal of Ω_i on Γ and $p \in \mathbb{R}$, $p > 0$ a constant.

It is possible to reduce algorithm (3)–(5) to the so-called interface variables. Define the operators $\mathscr{M}_i : (\lambda_i, f) \mapsto w_i$ solution of

$$\partial_t(\phi w_i) + \mathscr{L}w_i + \mathscr{F}(w_i) = q \quad \text{in } \Omega_i \times (0,T), \tag{6}$$

$$(\partial_{n_i} + p)w_i = \lambda_i \quad \text{on } \Gamma \times (0,T), \tag{7}$$

$$w_i^k(\cdot,0) = w_0 \text{ in } \Omega_i, \quad \mathscr{G}w_i^k = g \text{ on } \partial\Omega_i \setminus \Gamma \times (0,T). \tag{8}$$

Here $f = (q, w_0, g)$ represents all source terms except the ones on the interface Γ that are represented separately by λ_i. With these definitions, the transmission conditions (4) can be written as $\lambda_i^{k+1} = -\lambda_\times^k + 2p\mathscr{M}_\times(\lambda_\times^k, f)$, and as a system

$$\begin{pmatrix}\lambda_1^k\\ \lambda_2^k\end{pmatrix} = \begin{pmatrix}-\lambda_2^{k-1} + 2p\mathscr{M}_2(\lambda_2^{k-1}, f)\\ -\lambda_1^{k-1} + 2p\mathscr{M}_1(\lambda_1^{k-1}, f)\end{pmatrix}. \tag{9}$$

The SWR algorithm (3) is therefore a fixed point algorithm for the nonlinear interface problem

$$\begin{pmatrix}\lambda_1\\ \lambda_2\end{pmatrix} = \begin{pmatrix}-\lambda_2 + 2p\mathscr{M}_2(\lambda_2, f)\\ -\lambda_1 + 2p\mathscr{M}_1(\lambda_1, f)\end{pmatrix}. \tag{10}$$

Each iterate requires solving the nonlinear problem (6)–(8). This can be achieved by a Newton method, or a semi-implicit discretisation in time. The latter method has been implemented in [4] for the advection diffusion reaction equation, where the convergence of the fixed point algorithm has been proved. The extension of the proof to the system (1) should be easy.

4 Newton-Schwarz Optimised Waveform Relaxation

The new approach consists first in solving the system (10) by a Newton algorithm. If the interface problem is well-posed, and if the initial data for Newton is sufficiently closed to the solution, the algorithm converges to that solution. According to the interface problem (10), we seek the zeros of the nonlinear function

$$\Theta(\lambda) := -(\lambda_1 + \lambda_2)\begin{pmatrix}1\\ 1\end{pmatrix} + 2p\Upsilon(\lambda), \quad \Upsilon(\lambda) := \begin{pmatrix}\mathscr{M}_2(\lambda_2, f)\\ \mathscr{M}_1(\lambda_1, f)\end{pmatrix}.$$

One step $k-1 \to k$ of Newton's algorithm consists in solving the linear system $\Theta'(\lambda^{k-1}) \cdot (\lambda^k - \lambda^{k-1}) = -\Theta(\lambda^{k-1})$. To evaluate the derivative of Θ, we must calculate the derivative of the functions $\lambda_i \mapsto \mathscr{M}_i(\lambda_i, f)$. If $w_i = \mathscr{M}_i(\lambda_i, f)$ and $W_i = \mathscr{M}_i(\lambda_i + \tilde{\lambda}_i, f)$, we see by subtracting equations (6) for w_i and W_i, that $W_i - w_i$ is solution of

$$\partial_t(\phi(W_i - w_i)) + \mathscr{L}(W_i - w_i) + \mathscr{F}(W_i) - \mathscr{F}(w_i) = 0.$$

Introducing the derivative of $\mathscr{F}$, $\mathscr{F}(W_i) - \mathscr{F}(w_i) = \mathscr{F}'(w_i)(W_i - w_i) + \mathscr{O}((W_i - w_i)^2)$, and therefore $W_i - w_i = \tilde{w}_i + o(\tilde{w}_i^2)$, where $\tilde{w}_i$ is solution of the linear equation

$$\partial_t(\phi\tilde{w}_i) + \mathscr{L}\tilde{w}_i + \mathscr{F}'(w_i)\tilde{w}_i = 0. \tag{11}$$

$$(\partial_{n_i} + p)\tilde{w}_i = \tilde{\lambda}_i \tag{12}$$

$$\tilde{w}_i(x,0) = 0 \text{ in } \Omega_i, \quad \mathscr{G}\tilde{w}_i = 0 \text{ on } \partial\Omega_i \setminus \Gamma \times (0,T). \tag{13}$$

Therefore $\partial_{\lambda_i}\mathscr{M}_i(\lambda_i, f) \cdot \tilde{\lambda}_i = \tilde{w}_i := \mathscr{M}^{lin}(\mathscr{F}'(w_i); \tilde{\lambda}_i)$, and

$$\Theta'(\lambda) \cdot \tilde{\lambda} = -(\tilde{\lambda}_1 + \tilde{\lambda}_2)\begin{pmatrix}1\\ 1\end{pmatrix} + 2p\begin{pmatrix}\mathscr{M}^{lin}(\mathscr{F}'(w_2); \tilde{\lambda}_2)\\ \mathscr{M}^{lin}(\mathscr{F}'(w_1); \tilde{\lambda}_1)\end{pmatrix}.$$

After these computations, the algorithm takes the form

$$w_i^{k-1} = \mathscr{M}_i(\lambda_i^{k-1}, f),$$

$$-\sum_{i=1}^{2}(\lambda_i^k - \lambda_i^{k-1})\begin{pmatrix}1\\1\end{pmatrix} + 2p\begin{pmatrix}\mathscr{M}_2^{lin}(\mathscr{F}'(w_2^{k-1});\lambda_2^k - \lambda_2^{k-1})\\ \mathscr{M}_1^{lin}(\mathscr{F}'(w_1^{k-1});\lambda_1^k - \lambda_1^{k-1})\end{pmatrix} = \tag{14}$$
$$-\sum_{i=1}^{2}\lambda_i^k\begin{pmatrix}1\\1\end{pmatrix} + 2p\begin{pmatrix}\mathscr{M}_2(\lambda_2^{k-1}, f)\\ \mathscr{M}_1(\lambda_1^{k-1}, f)\end{pmatrix}$$

The approach requires in every iteration to solve two nonlinear problems in the subdomains. Therefore, a nested iterative procedure is necessary (Newton, or semi-implicit time stepping). Once this is done, $\lambda^{n+1} - \lambda^n$ is a solution of a linear problem solved in parallel in the subdomains.

5 Implementation Using Newton-Krylov Methods and Numerical Results

We have implemented both the classical and the new approach for a special case of problem (1). We assume that s = 2 and $w = (u, v)$ where u denotes a mobile species and v denotes a fixed species. The nonlinear function $\mathscr{F}$ is given by $\mathscr{F}(w) = (R(u,v), -R(u,v))$ where $R(u,v)$ is the overall reaction rate of the reversible reaction $u \leftrightarrows v$.

For the computation of $\mathscr{M}_i(\lambda_i^{k-1}, f)$, we use an implicit Euler scheme in time and a hybrid finite volume scheme (based on [7]) in space. The nonlinear systems are then treated with a global implicit approach by means of Newton's method with exact LU-decomposition. The linear interface problems (14) for λ_i^k are solved using GMRES as Krylov-type method with a precision strategy in the spirit of inexact Newton methods: we adapt the precision of the linear solver with respect to the residuals of the Newton iterates and save therefore costly subdomain evaluations.

Concerning the stopping criterion for the Newton-Schwarz optimised algorithm, it is classically controlled by both the residual and the correction ($\Delta\lambda$) norm. The Schwarz optimised algorithm is only controlled by the correction norm.

For all tests, we set the simulation domain to $\Omega = [0,1] \times [0,1] \subset \mathbb{R}^2$ with the subdomains $\Omega_1 = [0, 0.5] \times [0, 1]$ and $\Omega_2 = [0.5, 1] \times [0, 1]$. The time window considered is $t \in [0, 1]$. Physical parameters are $\phi = 1$, $a = 1.5$, $(b_x, b_y) = (5 \cdot 10^{-2}, 1 \cdot 10^{-3})$. The nonlinear coupling term is defined by $R(u,v) = k(v - \Psi(u))$ where the function Ψ is a BET isotherm law defined by

$$\Psi(u) = \frac{Q_s K_L u}{(1 + K_L u - K_S u)(1 - K_S u)}.$$

BET theory is a rule for the physical adsorption of gas molecules on a solid surface and serves as the basis for an important analysis technique for the measurement of the specific surface area of a material (cf. [3]). This law is insofar mathematically interesting as it is neither convex nor concave (cf. Fig. 1) and is therefore a challenging problem for standard nonlinear solvers like Newton's method. We set $k = 100$,

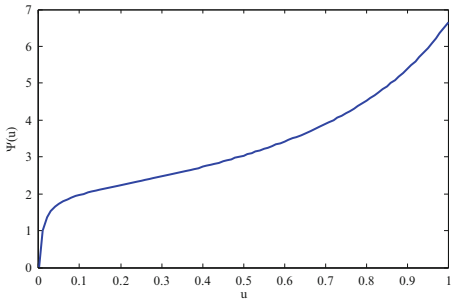


Fig. 1. BET Isotherm law function Ψ with $Q_S = 2$, $K_S = 0.7$, $K_L = 100$

$Q_S = 2$, $K_S = 0.7$ and $K_L = 100$. Initial values are set to $(u_0, v_0) = (\frac{1}{2}, \frac{1}{3})$. By defining the function $g(x,y,t) = (\sin(\pi x)\cos(\pi y)\cos(2\pi t) + \cos(\pi x)\sin(\pi y)\cos(2\pi t) + \cos(\pi x)\cos(\pi y)\sin(2\pi t) + 1)/2$ we impose Dirichlet boundary conditions with values set to $u(x,y,t) = g(x,y,t)$ for $(x,y) \in \partial\Omega$.

As a first experiment, we are interested in the sensitivity of the new approach with respect to the parameter p of the Robin transmission condition. Indeed the theory of optimised Schwarz waveform relaxation for linear problem relies on the fact that the convergence properties of the algorithm heavily depend on this parameter. A best parameter for the advection diffusion reaction equation can be found analytically by solving a best approximation problem, see [1, 8]. No such analysis is available for the nonlinear problem, it is therefore interesting to study the issue numerically.

We discretise the numerical domain with $\Delta x = \Delta y = 1/40$ and $\Delta t = 1/10$ and impose a random initial guess on the interface for the first iteration. As both subdomains are the same size, the number of overall matrix inversions is a meaningful criterion for measuring the numerical performance. We run the two approaches for different parameters p of the Robin transmission condition and plot in Fig. 2 (left) the number of matrix inversions as a function of the parameter p in the Robin transmission condition. One observes first that the performance of the classical approach depends highly on the parameter p of the Robin transmission condition, as in the linear case. The best parameter is $p^* \approx 40$. We observe that the new approach also shows the best performance at p^* but is much less sensitive to the choice of the parameter. The loss of sensitivity with respect to the parameter is still an open question.

It turns out that the new method has a cost overhead, that becomes non negligible if space discretisations are chosen too coarse. For this reason, we study the asymptotic behaviour of the two approaches using always the optimal parameter of the classical approach. We refine the problem in space using always $\Delta x = \Delta y$. Note that we keep the time step constant at $\Delta t = 0.1$. Refining the discretisation also in time would lead to a problem that is quasi stationary at every time step since we use a global implicit approach. We measure again the overall number of matrix inversions in the two approaches and plot them in Fig. 2 (right) versus the discretisation size. One observes that the overhead cost of the new approach compared to the classical approach becomes negligible starting at a discretisation with about 150 grid points

per dimension for the new method. For problems finer than the respective thresholds, the new approach is always faster than the classical approach with the best parameter for the transmission condition. Moreover, the finer the discretisation, the larger the problem, the more important the accelerating property of the new approach. Note that the new approaches has a slope of $O(N^{1/7})$ in the asymptotic behaviour which is considerably less than the slope of the classical approach which behaves like $O(N^{1/2.75})$. The slopes have been determined graphically, no theoretical justification is available. However, this plot shows that the method is much less dependent of the size of the problems than the classical one.

In order to exemplify the accelerating property of the new approach, we perform a simulation with $N_x = N_y = 200$ points in each dimension keeping the number of time steps constant and compare the convergence behaviour of the stopping criteria of the two methods. In Fig. 3 we plot the convergence criterion versus the number of matrix inversions. Note that, for a better comparison, we set the residual norm of the nonlinear interface problem evaluated at the initial guess for both methods at zero matrix inversions. The classical approach exhibit a linear convergence followed by a superlinear convergence, similar to the behaviour of the linear algorithm. We observe the quadratic convergence of the new approach, the characteristic feature of the Newton algorithm.

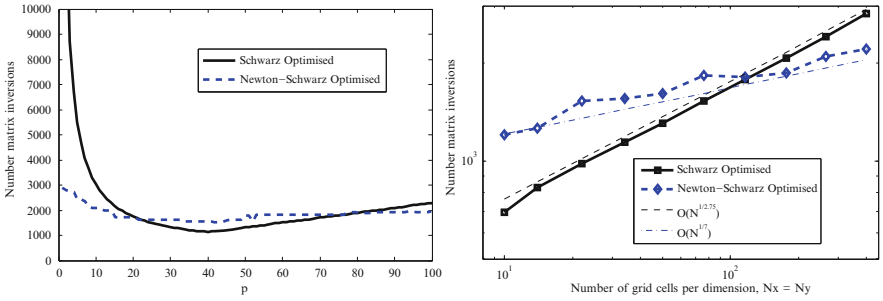


Fig. 2. Number of matrix inversions for the classical approach and new approach, synthetic test case. *Left*: Varying parameter p of the Robin transmission conditions with fixed discretisation in space and time. *Right*: Varying the number of discrete points per dimension ($N_x = N_y$) with fixed discretisation in time and optimal parameter for the Robin transmission condition

Finally, we want to apply the new approach to a benchmark test case in the context of CO_2 geological storage. The 3D test case is based on the benchmark for the SHPCO2 project (Simulation haute performance pour le stockage géologique du CO_2) which is described in [10]. The global domain is set to $\Omega = [0, 4{,}750] \times [0, 3{,}000] \times [-1{,}100, -1{,}000]$ with $(38, 24, 8)$ grid cells in (x, y, z)-direction. The domain is decomposed into the two nonoverlapping subdomains $\Omega_1 = [1{,}000, 2{,}500] \times [0, 3{,}000] \times [-1{,}050, -1{,}000]$ and $\Omega_2 = \Omega \setminus \Omega_1$. We call Ω_1 the reactive subdomain since in this subdomain an injection of the mobile species u is modelled by a source term. The initial state is zero for the mobile and immobile species. We consider

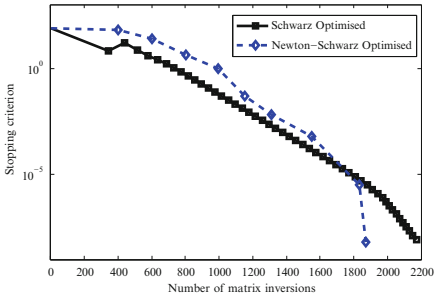


Fig. 3. Convergence history with 200 points per space dimension for the classical approach and new approach, synthetic test case

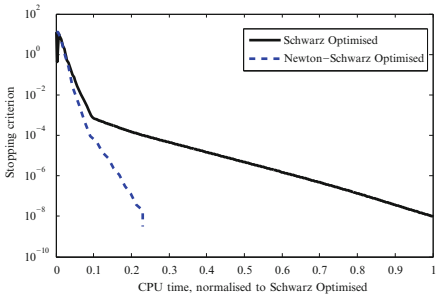


Fig. 4. Convergence history for the classical approach and new approach, SHPCO2 benchmark case

again the BET isotherm law as nonlinear coupling term. The injected mobile species is partially adsorbed by the reaction and partially transported by mainly advection. Simulation time is $[0, 100]$. The SWR approach allows us to use different discretisations in the subdomains. We choose to use ten time steps in the reactive subdomain Ω_1 and only five time steps in the subdomain Ω_2. This choice is insofar justified since the rapid injection in the reactive subdomain restricts the time step size by imposing a maximum number of Newton iterates of ten. As in the subdomain Ω_2, the mobile species appears only by transport processes on a slower time scale than the injection, one can choose a larger time step in order to respect the maximum number of Newton iterations. Concerning the parameter of the Robin transmission condition, we use a low frequency approximation of the optimal parameter. The initial guess on the interface is zero for both subdomain interfaces. In Fig. 4 we plot the convergence histogram, i.e. the stopping criterion in a logarithmic scale versus the CPU time (normalised to the CPU time of the classical approach). Note that both subdomains have a different size of unknowns and therefore the number of matrix inversions, as used in the previous examples, is no longer a valid tool to measure the effort. One observes that the new approach needs only about 20 % of the CPU time of the classical approach.

6 Conclusion

Based on a nonlinear coupled reactive transport system we have developed a new approach for solving the interface problem in the nonlinear case using Krylov-accelerators. In contrast to NKS methods the use of SWR methods allows us to use different time discretisations in the subdomains and so to localise time stepping constraints. We have implemented and tested the method, comparative results with the classical approach have been provided.

The numerical tests showed that, besides an overhead cost for coarse space discretisations, the method has an accelerating property and shows much less sensitivity with respect to the choice of the parameter of the Robin condition. The quadratic

convergence behaviour of the new approach outperforms the superlinear convergence behaviour of the classical approach. Nevertheless, the new approach does have significant overhead costs that are not negligible in the case of coarse problems. Note that a third approach is possible, namely to start with a Newton algorithm for the nonlinear problem, and to solve the so obtained linear problem by a Schwarz-Krylov algorithm (cf. [9]).

Bibliography

[1] D. Bennequin, M. Gander, and L. Halpern. A Homographic Best Approximation Problem with Application to Optimized Schwarz Waveform Relaxation. *Math. Comp.*, 78(265):185–223, 2009.

[2] E. Brakkee and P. Wilders. The Influence of Interface Conditions on Convergence of Krylov-Schwarz Domain Decomposition for the Advection-Diffusion Equation. *Journal of Scientific Computing*, 12:11–30, 1997.

[3] S. Brunauer, P. H. Emett, and E. Teller. Adsorption of Gases in Multimolecular Layers. *Journal American Chemical Society*, 60(2):309–319, 1938.

[4] F. Caetano, L. Halpern, M. Gander, and J. Szeftel. Schwarz waveform relaxation algorithms for semilinear reaction-diffusion. *Networks and heterogeneous media*, 5(3):487–505, 2010.

[5] X. C. Cai, W. D. Gropp, D. E. Keyes, and M. D. Tidriri. Parallel implicit methods for aerodynamics. In *In Keyes*, pages 465–470. American Mathematical Society, 1994.

[6] P. Cresta, O. Allix, C. Rey, and S. Guinard. Nonlinear localization strategies for domain decomposition methods: application to post-buckling analyses. *CMAME*, 196(8):1436–1446, 2007.

[7] R. Eymard, T. Gallouët, and R. Herbin. Discretization of heterogeneous and anisotropic diffusion problems on general nonconforming meshes sushi: a scheme using stabilization and hybrid interfaces. *IMA Journal of Numerical Analysis*, 30(4):1009–1043, 2010.

[8] M. J. Gander and L. Halpern. Optimized Schwarz waveform relaxation methods for advection reaction diffusion problems. *SIAM J. Numer. Anal.*, 45(2): 666–697, 2007.

[9] F. Haeberlein. *Time-Space Domain Decomposition Methods for Reactive Transport Applied to CO_2 Geological Storage*. PhD Thesis, University Paris 13, 2011.

[10] F. Haeberlein, A. Michel, and L. Halpern. A test case for multi-species reactive-transport in heterogeneous porous media applied to CO_2 geological storate. *http://www.ljll.math.upmc.fr/mcparis09/Files/haeberlein_poster.pdf*, 2009.

[11] J. Pebrel, C. Rey, and P. Gosselet. A nonlinear dual domain decomposition method: application to structural problems with damage. *international journal of multiscale computational engineering*, 6(3):251–262, 2008.

Alternating and Linearized Alternating Schwarz Methods for Equidistributing Grids

Martin J. Gander[1], Ronald D. Haynes[2], and Alexander J.M. Howse[2]

[1] Université de Genève Martin.Gander@unige.ch
[2] Memorial University of Newfoundland rhaynes@mun.ca

1 Introduction

The solution of partial differential equations (PDEs) with disparate space and time scales often benefit from the use of nonuniform meshes and adaptivity to successfully track local solution features.

In this paper we consider the problem of grid generation using the so–called equidistribution principle (EP) [3] and domain decomposition (DD) strategies. In the time dependent case, the EP is used to evolve an initial (often uniform) grid by **r**elocating a fixed number of mesh nodes. This leads to a class of adaptive methods known as **r**–refinement or moving mesh methods. A thorough recent review of moving mesh methods for PDEs can be found in the book [11].

In general, the appropriate grid for a particular problem depends on features of the (typically unknown) solution of the PDE. Here we will focus on the grid generation problem for the time independent, given function $u(x)$ of a single spatial variable $x \in [0,1]$. Given some positive measure $M(x)$ of the error or difficulty in the solution $u(x)$, the EP requires that the mesh points are chosen so that the error contribution on each interval $[x_{i-1}, x_i]$ is the same. The function M is known as the monitor or mesh density function. Mathematically, we may write this as

$$\int_{x_{i-1}}^{x_i} M(\tilde{x})\, d\tilde{x} \equiv \frac{1}{N}\int_0^1 M(\tilde{x}) d\tilde{x} \quad \text{or} \quad \int_0^{x(\xi_i)} M(\tilde{x})\, d\tilde{x} = \frac{i}{N}\theta \equiv \xi_i \theta, \qquad \text{(EP)}$$

where $x(\xi_i) = x_i$ and $\theta \equiv \int_0^1 M(\tilde{x})\, d\tilde{x}$ is the total error in the solution. The EP defines a one–to–one co-ordinate transformation between the physical co–ordinate x and underlying computational co–ordinate ξ. This will naturally concentrate mesh points where the error in the solution is large.

Differentiating the continuous formulation of EP gives the required mesh transformation, $x(\xi)$, as the solution of the nonlinear boundary value problem

$$\frac{d}{d\xi}\left\{ M(x(\xi))\frac{d}{d\xi}x(\xi) \right\} = 0, \quad x(0) = 0 \quad \text{and} \quad x(1) = 1. \qquad (1)$$

R. Bank et al. (eds.), *Domain Decomposition Methods in Science and Engineering XX*, 395
Lecture Notes in Computational Science and Engineering 91,
DOI 10.1007/978-3-642-35275-1_46, © Springer-Verlag Berlin Heidelberg 2013

If M is chosen properly, we expect the solution $u(x)$ to be easy to represent on a uniform grid in ξ. In general, the physical solution u is not known and instead satisfies a PDE. In that case, the mesh transformation, satisfying (1), and the physical PDE, are coupled and often solved in an iterative fashion.

We will assume (1) has a unique solution, see [8] for details. In [8], the authors consider the solution of (1) and time dependent extensions using classical parallel, optimized and optimal Schwarz methods. In this paper we continue the work of [8] by providing details of the nonlinear and linearized alternating Schwarz approaches. The reader is also referred to the experimental papers [7, 9, 10], which proposed various strategies to couple DD and moving meshes. See [1, 2, 4–6, 12–15] for a discussion of DD methods applied to other nonlinear PDEs.

In Sect. 2 we propose a new nonlinear alternating Schwarz method to solve (1) and prove convergence in L^∞. In Sect. 3 we avoid the nonlinear subdomain problems and propose and analyze a linearized alternating Schwarz algorithm. Brief numerical results are presented in the final section.

2 A Nonlinear Alternating Schwarz Method

In [8] we consider the solution of (1) by a parallel, classical nonlinear Schwarz iteration. On each subdomain a nonlinear BVP is solved and Dirichlet transmission conditions are used at the subdomain interfaces. Convergence of the iteration can be accelerated if we are willing to compute sequentially. Consider the nonlinear alternating Schwarz iteration

$$\begin{aligned} (M(x_1^n)x_{1,\xi}^n)_\xi = 0,\ \xi \in \Omega_1, \qquad & (M(x_2^n)x_{2,\xi}^n)_\xi = 0,\ \xi \in \Omega_2, \\ x_1^n(0) = 0, \qquad & x_2^n(\alpha) = x_1^n(\alpha), \\ x_1^n(\beta) = x_2^{n-1}(\beta), \qquad & x_2^n(1) = 1, \end{aligned} \tag{2}$$

where $\Omega_1 = (0,\beta)$ and $\Omega_2 = (\alpha,1)$ with $\alpha < \beta$.

Direct integration and enforcing the boundary conditions gives the following implicit representation of the subdomain solutions.

Lemma 1. *The subdomain solutions on Ω_1 and Ω_2 of (2) are given implicitly as*

$$\int_0^{x_1^n(\xi)} M(\tilde{x})\,d\tilde{x} = \frac{\xi}{\beta}\int_0^{x_2^{n-1}(\beta)} M(\tilde{x})\,d\tilde{x}, \tag{3}$$

and

$$\int_0^{x_2^n(\xi)} M(\tilde{x})\,d\tilde{x} = \frac{1-\xi}{1-\alpha}\int_0^{x_1^n(\alpha)} M(\tilde{x})\,d\tilde{x} + \frac{\xi-\alpha}{1-\alpha}\int_0^1 M(\tilde{x})\,d\tilde{x}. \tag{4}$$

Let $\|\cdot\|_\infty$ denote the usual L^∞ norm. We now relate $x_{1,2}^n$ to $x_{1,2}^{n-1}$ and obtain the following result.

Theorem 1. *Assume M is differentiable and there exist positive constants a and A satisfying $0 < a \le M(x) \le A < \infty$. Then the alternating Schwarz iteration (2) converges for any initial guess $x_2^0(\beta)$ and we have the error estimates*

$$||x - x_1^{n+1}||_\infty \le \rho^n \frac{A}{a} |x(\beta) - x_2^0(\beta)|, \quad ||x - x_2^{n+1}||_\infty \le \rho^n \frac{A}{a} |x(\alpha) - x_1^0(\alpha)|, \tag{5}$$

with contraction factor $\rho := \frac{\alpha}{\beta} \frac{1-\beta}{1-\alpha} < 1$.

Proof. Evaluating (3) at $\xi = \alpha$ and using the expression for $x_2^{n-1}(\beta)$ from (4) we have

$$\int_0^{x_1^n(\alpha)} M \, d\tilde{x} = \frac{\alpha}{\beta} \left\{ \frac{\beta - 1}{\alpha - 1} \int_0^{x_1^{n-1}(\alpha)} M \, d\tilde{x} + \frac{\beta - \alpha}{1 - \alpha} \int_0^1 M \, d\tilde{x} \right\}.$$

Defining the two quantities

$$K_1^n = \int_0^{x_1^n(\alpha)} M(\tilde{x}) \, d\tilde{x} \qquad \text{and} \qquad C = \int_0^1 M(\tilde{x}) \, d\tilde{x},$$

we obtain the linear iteration

$$K_1^n = \frac{\alpha}{\beta} \frac{\beta - 1}{\alpha - 1} K_1^{n-1} + \frac{\alpha}{\beta} \frac{\beta - \alpha}{1 - \alpha} C. \tag{6}$$

This iteration converges with rate $\rho := \frac{\alpha}{\beta} \frac{1-\beta}{1-\alpha} < 1$, and has the limit

$$K_1^* = \frac{\alpha}{\beta} \frac{1 - \beta}{1 - \alpha} K_1^* + \frac{\alpha}{\beta} \frac{\beta - \alpha}{1 - \alpha} C \quad \Longrightarrow K_1^* = \alpha C. \tag{7}$$

Since the monodomain solution also satisfies

$$\int_0^{x(\alpha)} M(\tilde{x}) \, d\tilde{x} = \alpha C,$$

and $M(x) \ge a > 0$, we have convergence at the interface to the correct limit.

Subtracting (6) from (7) we have

$$\int_{x_1^n(\alpha)}^{x(\alpha)} M(\tilde{x}) \, d\tilde{x} = \rho^n \int_{x_1^0(\alpha)}^{x(\alpha)} M(\tilde{x}) \, d\tilde{x}. \tag{8}$$

Subtracting (4) from the equivalent expression for the exact solution and using (8) we obtain

$$\int_{x_2^{n+1}(\xi)}^{x(\xi)} M(\tilde{x}) \, d\tilde{x} = \frac{1 - \xi}{1 - \alpha} \int_{x_1^n(\alpha)}^{x(\alpha)} M(\tilde{x}) \, d\tilde{x} = \frac{1 - \xi}{1 - \alpha} \rho^n \int_{x_1^0(\alpha)}^{x(\alpha)} M(\tilde{x}) \, d\tilde{x}.$$

Taking the modulus and using the boundedness of M we obtain, for all $\xi \in [\alpha, 1]$,

$$|x(\xi) - x_2^{n+1}(\xi)| \le \frac{1 - \xi}{1 - \alpha} \rho^n \frac{A}{a} |x(\alpha) - x_1^0(\alpha)|.$$

Taking the supremum gives the second estimate in (5). The estimate on subdomain one is obtained similarly. □

3 A Linearized Alternating Schwarz Method

We may avoid nonlinear solves on each subdomain in (2) by considering a linearized alternating Schwarz iteration,

$$\begin{aligned} (M(x_1^{n-1})x_{1,\xi}^n)_\xi = 0,\ \xi \in \Omega_1 \qquad & (M(x_2^{n-1})x_{2,\xi}^n)_\xi = 0,\ \xi \in \Omega_2 \\ x_1^n(0) = 0, \qquad & x_2^n(\alpha) = x_1^n(\alpha), \\ x_1^n(\beta) = x_2^{n-1}(\beta), \qquad & x_2^n(1) = 1. \end{aligned} \tag{9}$$

At iteration n we evaluate the nonlinear diffusion coefficient M using the solution obtained from the previous iterate and obtain the updated solution by a single linear BVP solve on each subdomain. A simple calculation yields the following representation of the subdomain solutions.

Lemma 2. *The subdomain solutions of (9) are given by*

$$x_1^n(\xi) = x_2^{n-1}(\beta)\frac{\int_0^\xi \frac{d\tilde{\xi}}{M(x_1^{n-1}(\tilde{\xi}))}}{\int_0^\beta \frac{d\tilde{\xi}}{M(x_1^{n-1}(\tilde{\xi}))}}, \tag{10}$$

and

$$x_2^n(\xi) = x_1^n(\alpha) + (1 - x_1^n(\alpha))\frac{\int_\alpha^\xi \frac{d\tilde{\xi}}{M(x_2^{n-1}(\tilde{\xi}))}}{\int_\alpha^1 \frac{d\tilde{\xi}}{M(x_2^{n-1}(\tilde{\xi}))}}. \tag{11}$$

Convergence of the linearized alternating Schwarz iteration (9) follows by proving convergence at the interior interfaces and showing we have converged to the correct limit.

Theorem 2. *Under the assumptions of Theorem 1, the linearized alternating Schwarz iteration (9) converges for any smooth initial guesses $x_1^0(\xi)$ and $x_2^0(\xi)$.*

Proof. Evaluating the subdomain solutions (10) and (11) at the interfaces, we obtain for the interface values the iterations

$$x_1^n(\alpha) = \mathscr{C}_\alpha^n x_1^{n-1}(\alpha) + \mathscr{D}_\alpha^n \quad \text{and} \quad x_2^n(\beta) = \mathscr{C}_\beta^n x_2^{n-1}(\beta) + \mathscr{D}_\beta^n,$$

where

$$\mathscr{C}_\alpha^n = \frac{\int_\beta^1 \frac{d\tilde{\xi}}{M(x_2^{n-2}(\tilde{\xi}))}}{\int_\alpha^1 \frac{d\tilde{\xi}}{M(x_2^{n-2}(\tilde{\xi}))}} \frac{\int_0^\alpha \frac{d\tilde{\xi}}{M(x_1^{n-1}(\tilde{\xi}))}}{\int_0^\beta \frac{d\tilde{\xi}}{M(x_1^{n-1}(\tilde{\xi}))}}, \qquad \mathscr{D}_\alpha^n = \frac{\int_\alpha^\beta \frac{d\tilde{\xi}}{M(x_2^{n-2}(\tilde{\xi}))}}{\int_\alpha^1 \frac{d\tilde{\xi}}{M(x_2^{n-2}(\tilde{\xi}))}} \frac{\int_0^\alpha \frac{d\tilde{\xi}}{M(x_1^{n-1}(\tilde{\xi}))}}{\int_0^\beta \frac{d\tilde{\xi}}{M(x_1^{n-1}(\tilde{\xi}))}},$$

and

$$\mathscr{C}_\beta^n = \frac{\int_\beta^1 \frac{d\tilde{\xi}}{M(x_2^{n-1}(\tilde{\xi}))}}{\int_\alpha^1 \frac{d\tilde{\xi}}{M(x_2^{n-1}(\tilde{\xi}))}} \frac{\int_0^\alpha \frac{d\tilde{\xi}}{M(x_1^{n-1}(\tilde{\xi}))}}{\int_0^\beta \frac{d\tilde{\xi}}{M(x_1^{n-1}(\tilde{\xi}))}}, \qquad \mathscr{D}_\beta^n = \frac{\int_\alpha^\beta \frac{d\tilde{\xi}}{M(x_2^{n-1}(\tilde{\xi}))}}{\int_\alpha^1 \frac{d\tilde{\xi}}{M(x_2^{n-1}(\tilde{\xi}))}}.$$

It is possible to show the quantities $\mathscr{C}_\alpha^n, \mathscr{D}_\alpha^n, \mathscr{C}_\beta^n$ and $\mathscr{D}_\beta^n$ satisfy

$$0 < \mathscr{C}_\alpha^n, \mathscr{C}_\beta^n \le \rho < 1, \quad 0 < \mathscr{D}_\alpha^n \le D_\alpha < 1, \quad \text{and} \quad 0 < \mathscr{D}_\beta^n \le D_\beta < 1,$$

where

$$\rho := \frac{1}{1+\frac{a}{A}\frac{\beta-\alpha}{1-\beta}} \frac{1}{1+\frac{a}{A}\frac{\beta-\alpha}{\alpha}}, \quad D_\alpha := \frac{1}{1+\frac{a}{A}\frac{\beta-\alpha}{\alpha}} \frac{1}{1+\frac{a}{A}\frac{1-\beta}{\beta-\alpha}}, \quad \text{and} \quad D_\beta := \frac{1}{1+\frac{a}{A}\frac{1-\beta}{\beta-\alpha}}.$$

To establish these bounds let $F(x) := 1/M(x)$. The assumptions on M imply $\frac{1}{A} \le F(x) \le \frac{1}{a}$. As an example, the upper and lower bounds on F then imply

$$\frac{\int_0^\alpha F(x(\xi))\,d\xi}{\int_0^\beta F(x(\xi))\,d\xi} \le \frac{1}{1+\frac{a}{A}\frac{\beta-\alpha}{\alpha}} \quad \text{and} \quad \frac{\int_\beta^1 F(x(\xi))\,d\xi}{\int_\alpha^1 F(x(\xi))\,d\xi} \le \frac{1}{1+\frac{a}{A}\frac{\beta-\alpha}{1-\beta}}.$$

Consider now the iteration for $x_1^n(\alpha)$ only. Using the recursion, we have

$$x_1^n(\alpha) = \prod_{k=1}^{n} \mathscr{C}_\alpha^k x_1^0(\alpha) + \sum_{k=1}^{n} \mathscr{D}_\alpha^k \left(\prod_{l=k+1}^{n} \mathscr{C}_\alpha^l \right),$$

where the product in the k–th term of the sum is assumed to be one if the lower index of the product exceeds the upper index. Since $\rho < 1$, the product multiplying $x_1^0(\alpha)$ must go to zero as $n \to \infty$. The infinite series converges by direct comparison with $\sum_{k=1}^{\infty} D_\alpha \rho^{k-1}$. A corresponding argument applies to show convergence of $x_2^n(\beta)$.

Denote the limits of $\{x_1^n(\alpha)\}$ and $\{x_2^n(\beta)\}$ as $\tilde{x}_\alpha$ and $\tilde{x}_\beta$ respectively. Since the interface values converge, the subdomain solutions defined by (9) converge to functions $\tilde{x}_1$ and $\tilde{x}_2$ both satisfying the nonlinear PDE. Since $\tilde{x}_1(\alpha) = \tilde{x}_2(\alpha)$ and $\tilde{x}_1(\beta) = \tilde{x}_2(\beta)$, both $\tilde{x}_1$ and $\tilde{x}_2$ satisfy the same PDE in the overlap with the same two boundary conditions, and by assumption of uniqueness, $\tilde{x}_1$ and $\tilde{x}_2$ must coincide in the overlap. One can therefore simply glue these two solutions together in order to obtain a function which satisfies the PDE everywhere, and also the two original boundary conditions at 0 and 1. Again by uniqueness, this must now be the desired solution. □

4 Numerical Results

In this section we numerically demonstrate the results above using a simple finite difference discretization of the BVP (1) and iterations (2) and (9). We also include results using nonlinear and linearized parallel Schwarz algorithm from [8] for comparison. Details of the numerical approach and convergence of the discrete DD algorithm will be considered elsewhere.

We solve EP for $u(x) = (1-e^{\lambda x})/(1-e^{\lambda})$ on the interval $x \in [0,1]$. For large values of λ this function exhibits a boundary layer at $x = 1$. We use the arc–length monitor function $M(x,u(x)) = \sqrt{1+u_x^2}$ and choose $\lambda = 20$. The errors reported in Figs. 1 and 2 are the differences between the single domain numerical solution and the domain decomposition solution over the first subdomain.

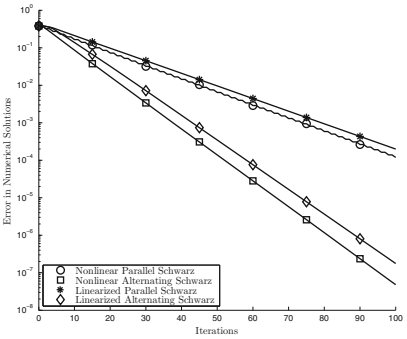


Fig. 1. Error versus # of DD iterations

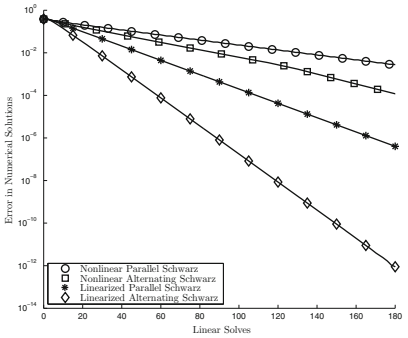


Fig. 2. Error versus # of linear solves

In Fig. 1 we solve (1) on two subdomains with a 5 % overlap using linearized and nonlinear, parallel and alternating Schwarz iterations. We see that the convergence of the alternating iteration is faster than the parallel algorithms for both the nonlinear and linearized versions of the algorithms. In terms of number of iterations the nonlinear algorithms outperform the linearized variants. It is important, however, to keep in mind that each nonlinear DD iteration is more expensive than its linearized counterpart. In Fig. 2 we repeat the convergence history as a function of a *work unit* which we take to be the cost of a linear solve. Each iteration of a linearized Schwarz algorithm requires one linear solve while each iteration of a nonlinear Schwarz algorithm requires many linear solves – one for each Newton step. Each linear solve required by both algorithms has roughly the same cost due to the structure of the Jacobian matrix. As a function of the work effort the efficacy of the linearized Schwarz algorithms is obvious for this example.

In Table 1 we demonstrate the quality of the computed grids by calculating the $\|\cdot\|_\infty$ error between $u(x)$ and the piecewise linear interpolant for $u(x)$ on grids obtained by the nonlinear and linearized alternating Schwarz algorithms, as a function of the number of iterations. The last column shows the interpolation error obtained with the single domain grid: the solution of (1) computed on a uniform ξ grid consisting of 101 points. All interpolation errors are computed using a very fine grid. The results show that the nonlinear Schwarz method is quickly able to find an appropriate grid transformation after a few DD iterations. The linearized Schwarz algorithm, as expected, requires more DD iterations but is able to find a quality grid efficiently due to the smaller relative cost per iteration.

Iterations	1	3	5	7	9	11	∞
Nonlinear	0.3625	0.0498	0.0462	0.0436	0.0449	0.0517	0.0366
Linearized	0.3625	0.1290	0.1019	0.0625	0.0453	0.0435	0.0366

Table 1. Interpolation errors for the grids obtained by Schwarz iterations.

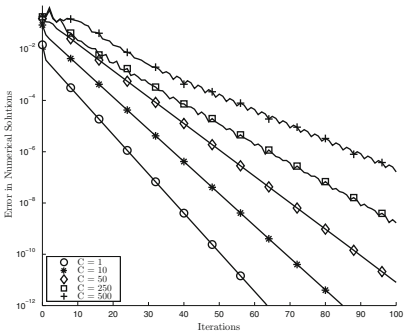


Fig. 3. Linearized Schwarz: error for varying C

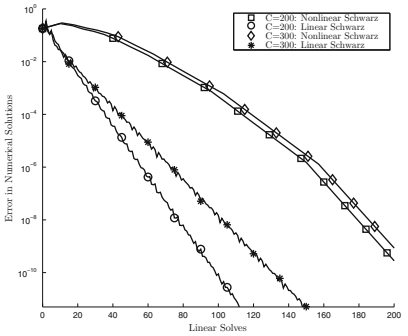


Fig. 4. Non-linear versus linearized Schwarz with varying C

The quantities ρ, D_α and D_β corresponding to iteration (9) and the error estimates in Theorem 1 indicate a dependence on the shape of M for the linearized alternating Schwarz iteration. To test this effect, we consider the performance of (9) for $M(x) = C(x-0.5)^2+1$. The parameter C controls the ratio a/A. As $C \to \infty$, $a/A \to 0$, and the contraction rate could diminish. This is demonstrated in Fig. 3. Figure 4 illustrates the effect of changing the value of C on both the nonlinear and linearized Schwarz algorithms. We see that the linearized Schwarz algorithm is affected more by an increase in C.

In summary, we have proposed, analyzed and provided brief numerical comparisons for two alternating Schwarz algorithms to solve the steady grid generation problem using the EP. Ongoing work includes the analysis of DD approaches to moving mesh PDEs for the time dependent mesh generation problem, the discrete analysis and extensions to higher dimensions.

Bibliography

[1] Igor P. Boglaev. Iterative algorithms of domain decomposition for the solution of a quasilinear elliptic problem. *J. Comput. Appl. Math.*, 80(2):299–316, 1997.

[2] Xiao-Chuan Cai and Maksymilian Dryja. Domain decomposition methods for monotone nonlinear elliptic problems. In *Domain decomposition methods in scientific and engineering computing (University Park, PA, 1993)*, volume 180 of *Contemp. Math.*, pages 21–27. Amer. Math. Soc., Providence, RI, 1994.

[3] C. de Boor. Good approximation by splines with variable knots. II. In *Conference on the Numerical Solution of Differential Equations (Univ. Dundee, Dundee, 1973)*, pages 12–20. Lecture Notes in Math., Vol. 363. Springer, Berlin, 1974.

[4] M. Dryja and W. Hackbusch. On the nonlinear domain decomposition method. *BIT*, 37(2):296–311, 1997.

[5] Martin J. Gander. A waveform relaxation algorithm with overlapping splitting for reaction diffusion equations. *Numerical Linear Algebra with Applications*, 6:125–145, 1998.

[6] Martin J. Gander and Christian Rohde. Overlapping Schwarz waveform relaxation for convection dominated nonlinear conservation laws. *SIAM J. Sci. Comp.*, 27(2):415–439, 2005.

[7] Ronald D. Haynes. Recent advances in Schwarz waveform relaxation moving mesh methods – a new moving subdomain method. In *Domain decomposition methods in science and engineering XIX*, volume 78 of *Lect. Notes Comput. Sci. Eng.*, pages 253–260. Springer, Berlin, 2011.

[8] Ronald D. Haynes and Martin J. Gander. Domain decomposition approaches for mesh generation via the equidistribution principle, 2011. Preprint.

[9] Ronald D. Haynes, Weizhang Huang, and Robert D. Russell. A moving mesh method for time-dependent problems based on Schwarz waveform relaxation. In *Domain decomposition methods in science and engineering XVII*, volume 60 of *Lect. Notes Comput. Sci. Eng.*, pages 229–236. Springer, Berlin, 2008.

[10] Ronald D. Haynes and Robert D. Russell. A Schwarz waveform moving mesh method. *SIAM J. Sci. Comput.*, 29(2):656–673, 2007.

[11] Weizhang Huang and Robert D. Russell. *Adaptive Moving Mesh Methods*, volume 174 of *Applied Mathematical Sciences*. Springer-Verlag, 2011.

[12] P.-L. Lions. On the Schwarz alternating method. I. In *First International Symposium on Domain Decomposition Methods for Partial Differential Equations (Paris, 1987)*, pages 1–42. SIAM, Philadelphia, PA, 1988.

[13] S. H. Lui. On monotone and Schwarz alternating methods for nonlinear elliptic PDEs. *M2AN Math. Model. Numer. Anal.*, 35(1):1–15, 2001.

[14] S. H. Lui. On linear monotone iteration and Schwarz methods for nonlinear elliptic PDEs. *Numer. Math.*, 93(1):109–129, 2002.

[15] Xue-Cheng Tai and Magne Espedal. Rate of convergence of some space decomposition methods for linear and nonlinear problems. *SIAM J. Numer. Anal.*, 35(4):1558–1570 (electronic), 1998.

Stability Analysis of the Matrix-Free Linearly Implicit Euler Method

Adrian Sandu[1] and Amik St-Cyr[2]

[1] Computational Science Laboratory, Department of Computer Science, Virginia Polytechnic Institute, Blacksburg, VA, 24060, USA. E-mail sandu@cs.vt.edu
[2] National Center for Atmospheric Research, 1850 Table Mesa Drive, Boulder CO, 80305. E-mail amik@ucar.edu.

Summary. Implicit time stepping methods are useful for the simulation of large scale PDE systems because they avoid the time step limitations imposed by explicit stability conditions. To alleviate the challenges posed by computational and memory constraints, many applications solve the resulting linear systems by iterative methods where the Jacobian-vector products are approximated by finite differences. This paper explains the relation between a linearly implicit Euler method, solved using a Jacobian-free Krylov method, and explicit Runge-Kutta methods. The case with preconditioning is equivalent to a Rosenbrock-W method where the approximate Jacobian, inverted at each stage, corresponds directly to the preconditioner. The accuracy of the resulting Runge-Kutta methods can be controlled by constraining the Krylov solution. Numerical experiments confirm the theoretical findings.

1 Introduction

Large systems of time dependent partial differential equations (PDEs), arising in multi-physics simulations, are often discretized using the method of lines approach. The independent time and space numerical schemes allow the coupling of multiple physics modules, and provide maximum flexibility in choosing appropriate algorithms. After the semi-discretization in space the system of PDEs is reduced to a system of ordinary differential equations (ODEs)

$$y' = f(y), \quad t_0 \le t \le t_f, \quad y(t_0) = y_0. \tag{1}$$

Here $y(t) \in \mathbb{R}^d$ is the solution vector and y_0 the initial condition. We denote the Jacobian of the ODE function by $J(y) = f_y(y) \in \mathbb{R}^{d\times d}$, and the identity matrix by $\mathbb{I} \in \mathbb{R}^{d\times d}$.

Stability requirements (e.g., the CFL condition for discretized hyperbolic PDEs) limit the time steps allowable by explicit time discretizations of (1). When the fastest time scales in the system (1) are short, e.g., in the presence of fast waves, the stability condition imposes time steps much smaller then those required to achieve the target

R. Bank et al. (eds.), *Domain Decomposition Methods in Science and Engineering XX*, Lecture Notes in Computational Science and Engineering 91,
DOI 10.1007/978-3-642-35275-1_47,

accuracy. The step size limitation by linear stability conditions is referred to as stiffness. In order to overcome this computational inefficiency, it is desirable to use implicit, unconditionally stable discretizations which allow arbitrarily large time steps [2]. Implicit methods have a high cost per step due to the need to solve a (non)linear system of equations.

To reduce the computational and memory costs of direct linear system solvers, and to aid parallelization, iterative Krylov space methods are employed. Furthermore, matrix-free implementations approximate Jacobian vector products by finite differences [4]. This approach avoids additional coding for the Jacobian, preserves the parallel scalability of the explicit model, and has become popular in many applications, e.g., [1, 5, 6]. The hope is that the properties of the implicit time discretization remain unaltered, provided that the iterative solutions are carried out to sufficient accuracy. *We show here that the matrix-free approach does alter the properties of the underlying implicit time stepping method.*

This study treats a linearly implicit method, together with the Krylov subspace iterations for solving the linear system, as a single numerical scheme. The analysis reveals that matrix-free implementations of linearly implicit methods are equivalent to explicit Runge Kutta methods. Consequently, the unconditional stability property of the base method is lost. When preconditioning is used, the matrix-free implicit methods are equivalent to Rosenbrock-W (ROS-W) methods where the approximate Jacobians correspond directly to the preconditioners.

2 The Matrix-Free Linearly Implicit Euler Method

Consider the linearly implicit Euler (LIE) method applied to (1)

$$\left(\mathbb{I} - \Delta t\, J(y_n)\right) \cdot w = f(y_n), \quad y_{n+1} = y_n + \Delta t \cdot w. \tag{2}$$

When the linear system is solved exactly (modulo roundoff errors) by LU factorization the method (2) is unconditionally stable, and thus suitable for the solution of stiff systems. For many PDEs semi-discretized in the method of lines framework, however, the dimension of the linear system (2) is very large, and the computational and memory costs associated with a direct solution are prohibitive. Moreover, the construction of the explicit Jacobian matrix J is difficult when the space discretization is based on a domain decomposition approach. To alleviate these problems, a popular approach is to solve (2) by matrix-free iterative methods. We seek to analyze the impact that this approximate solutions have on the stability and accuracy of the implicit time stepping scheme. *Our approach is to treat the original discretization (2) together with the iterations as a single numerical method applied to solve the ODE (1).*

To be specific, we solve the linear system in (2) by a Krylov space method. The initial guess is $y_{n+1} = y_n$, i.e., $w = 0$. After m iterations the following m-dimensional Krylov space is built:

$$\mathscr{K}_m = \text{span}\left\{ f(y_n), \ldots, \left(\mathbb{I} - \Delta t\, J(y_n)\right)^{m-1} f(y_n) \right\}.$$

In the matrix-free approach, the basis is constructed recursively and the Jacobian-vector products are approximated by finite differences

$$\ell_i = \ell_{i-1} - \Delta t\, \varepsilon^{-1}\, f(y_n + \varepsilon\, \ell_{i-1}) + \Delta t\, \varepsilon^{-1}\, \ell_1\,, \quad i = 2, \ldots, m\,. \tag{3}$$

We assume that the same scaling factor ε is used to compute the finite differences in all iterations. (The analysis can be easily extended to the case where a different ε is used in each iteration.) Denote

$$k_1 = f(y_n)\,; \quad k_i = f\left(y_n + \varepsilon \ell_{i-1}\right)\,, \ \ i = 2, \cdots, m\,. \tag{4}$$

The recurrence (3) can be expressed in terms of k_i as:

$$k_i = f\left(y_n + \Delta t\, \left(\Delta t^{-1}\, \varepsilon + (i-2)\right) k_1 - \Delta t \sum_{j=2}^{i-1} k_j\right), \ \ i = 2, \ldots, m. \tag{5}$$

The solution $w = \sum_{i=1}^{m} \alpha_i\, \ell_i \in \mathscr{K}_m$ can be expressed in terms of k_i's:

$$w = \left(\sum_{i=1}^{m} \alpha_i + \Delta t\, \varepsilon^{-1} \sum_{i=2}^{m} (i-1)\, \alpha_i\right) k_1 - \Delta t\, \varepsilon^{-1} \sum_{i=2}^{m} \left(\sum_{j=i}^{m} \alpha_j\right) k_i\,. \tag{6}$$

Equations (5) and (6), together with the relation $y_{n+1} = y_n + \Delta t\, w$, are compared with the m-stage explicit Runge Kutta (ERK) method [3]

$$k_i = f\left(y_n + \sum_{j=1}^{i-1} a_{ij}\, k_j\right), \ i = 1, \ldots, m\,; \quad y_{n+1} = y_n + \Delta t \sum_{i=1}^{m} b_i\, k_i\,.$$

The comparison reveals the following.

Theorem 1. *The matrix-free LIE (2) method is equivalent to an explicit Runge Kutta method. The number m of Krylov iterations defines the number of Runge Kutta stages.*

Equations (5) and (6) define the coefficients of the ERK method:

$$a_{i,1} = \Delta t^{-1}\, \varepsilon + (i-2)\,; \ \ a_{i,j} = -1\,, \ \text{ for } \ i = 2, \cdots, m\,, \ \ j = 2, \cdots, i-1\,;$$

$$b_1 = \sum_{j=1}^{m} \alpha_j + \Delta t\, \varepsilon^{-1} \sum_{j=2}^{m} (j-1)\, \alpha_j\,; \quad b_i = -\Delta t\, \varepsilon^{-1} \sum_{j=i}^{m} \alpha_j\,, \quad i = 2, \ldots, m\,.$$

2.1 Stability Considerations

The solution of the linear system (under the initial guess $w = 0$) is part of the Krylov space $\mathscr{K}_m$ and can be represented by a matrix polynomial

$$w = p_{m-1}\left(\mathbb{I} - \Delta t\, J(y_n)\right) \cdot f(y_n)\,.$$

The matrix-free LIE method applied to the Dahlquist test problem $y' = \lambda\, y$, $y(0) = 1$, gives the following solution:

$$y_{n+1} = y_n + \Delta t\, w = \left(1 + z\, p_{m-1}\,(1-z)\right)\, y_n = R(z)\, y_n\,,$$

with $z = \Delta t\, \lambda$. The stability function of the equivalent ERK method is the degree m polynomial $R(z) = 1 + z\, p_{m-1}\,(1-z)$.

Theorem 2. *The stability region of the LIE method, with a Krylov matrix-free linear solver, is necessarily finite. The unconditional stability of the original LIE method is lost.*

Similar considerations hold for Krylov space methods that use an orthogonal basis of the Krylov space, built by Arnoldi iterations [7].

2.2 Accuracy Considerations

The method accuracy is difficult to assess, as the coefficients depend on the time step. The relation between the finite difference scaling factor ε and the time step Δt is important in determining accuracy.

Assume that the finite difference scaling factor is a constant fraction of the time step, $\varepsilon/\Delta t = \text{const}$. This is a reasonable assumption: in order to increase accuracy one decreases both Δt, to reduce the truncation error, and ε, to reduce the finite difference error. (Of course, for very small ε the finite difference error becomes again large due to roundoff.) Also assume that the coefficients $\alpha_1, \ldots, \alpha_m$ do not depend on ε or Δt.

In this case the accuracy can be assessed using the classical approach. The order conditions depend on the Krylov space coefficients α as follows:

$$\text{Order 1:} \quad \sum_{i=1}^{m} b_i = \sum_{j=1}^{m} \alpha_j = 1\,, \tag{7a}$$

$$\text{Order 2:} \quad \sum_{i=1}^{m} b_i c_i = -\sum_{i=2}^{m} (i-1)\,\alpha_i = \frac{1}{2}\,. \tag{7b}$$

Neither condition (7a) nor (7b) are automatically satisfied by the Krylov iterative methods. In particular,

Lemma 1. *The first order accuracy of the matrix-free LIE is not automatic when $\varepsilon/\Delta t = const$. Additional constraints need to be imposed on the Krylov solution coefficients.*

Consider now the case where ε is constant (does not depend on Δt). Assume that the coefficients $\alpha_1, \ldots, \alpha_m$ do not depend on ε or Δt. A necessary condition for the method to be accurate of order p is that its stability function approximates the exponential, $R(z) = e^z + \mathcal{O}\left(z^{p+1}\right)$. The stability function does not depend on either ε or Δt. The conditions (7a) and (7b) on the Krylov solution coefficients $\alpha_1, \ldots, \alpha_m$, which are sufficient when $\varepsilon = const \cdot \Delta t$, seem to be necessary in the case $\varepsilon = const$.

In the general case the Krylov solution coefficients $\alpha_1, \ldots, \alpha_m$ do depend on Δt. For $\Delta t \to 0$ we have that $w \to f(y_n)$ and therefore $\alpha_1 \to 1$, $\alpha_2, \alpha_3, \ldots \to 0$. Asymptotically the condition (7a) holds. Moreover, the number of iterations m also depends on Δt through the convergence speed. Consequently, it is difficult to extend the classical accuracy analysis to matrix-free linearly implicit methods. It seems reasonable, however, to modify the Krylov method and impose at least condition (7a) on the Krylov coefficients.

3 Preconditioned Iterations

Consider the case where a preconditioner matrix M is used to speed up the iterations. The linear system (2) becomes

$$M^{-1}\left(\mathbb{I}-\Delta t\, J(y_n)\right)\cdot k = M^{-1} f(y_n)\,.$$

The Krylov space constructed in this case is

$$\mathscr{K}_m = \text{span}\left\{ f(y_n)\ \ \ldots,\ \left(M^{-1}\left(\mathbb{I}-\Delta t\, J(y_n)\right)\right)^{m-1} M^{-1} f(y_n)\right\}.$$

In the matrix-free approach the following basis is constructed recursively

$$\begin{aligned}
&\ell_1 = M^{-1} f(y_n)\,,\\
&\ell_i = M^{-1}\,\ell_{i-1} - \Delta t\;\varepsilon^{-1}\,M^{-1} f\left(y_n+\varepsilon\,\ell_{i-1}\right) + \Delta t\;\varepsilon^{-1}\,\ell_1\,,\quad i=2,\ldots,m\,.
\end{aligned}$$

Denote $k_1 = \Delta t\;\ell_1$ and $k_i = \Delta t\;\ell_1 - \varepsilon\,\ell_i$ for $i = 2,\cdots,m$. We have

$$\begin{aligned}
M\,k_1 &= \Delta t\; f(y_n)\\
M\,k_i &= \Delta t\; f\left(y_n + k_1 - k_{i-1}\right) + k_{i-1} - k_1\,,\quad i=2,\ldots,m\,.
\end{aligned} \tag{8}$$

Consider, for comparison, a Rosenbrock-W (ROW) method in the implementation-friendly formulation [2, Sect. IV.7]

$$\begin{aligned}
\left[\mathbb{I}-\Delta t\;\gamma\widehat{J}_n\right] k_i &= \Delta t\;\gamma f\left(y_n + \sum_{j=1}^{i-1} a_{ij}\,k_j\right) + \gamma\sum_{j=1}^{i-1} c_{ij}\,k_j\,,\\
y_{n+1} &= y_n + \sum_{i=1}^{s} m_i\,k_i\,.
\end{aligned} \tag{9}$$

Here $\widehat{J}_n \approx J(y_n)$ is an approximation of the exact Jacobian at the current step. We identify the method coefficients $\gamma = 1$ and

$$c_{i,1} = -1\,;\ \ c_{i,i-1} = 1\,;\ \ a_{i,1} = 1\,;\ \ a_{i,i-1} = -1\,,\quad i=2,\cdots,m\,.$$

From the solution $w = \sum_{i=1}^{m}\alpha_i\,\ell_i = \sum_{i=1}^{m} b_i\,k_i \in \mathscr{K}_m$ we identify the weights

$$b_1 = \alpha_1\,\Delta t^{-1} + \varepsilon^{-1}\sum_{j=2}^{m}\alpha_j\,;\quad b_i = -\varepsilon^{-1}\,\alpha_i\,,\ \ i=2,\ldots,m\,.$$

The preconditioner defines the Jacobian approximation in the ROW method,

$$M = \mathbb{I}-\Delta t\;\gamma\widehat{J}_n \quad\Rightarrow\quad \widehat{J}_n = \Delta t^{-1}\left(\mathbb{I}-M\right).$$

Theorem 3. *The preconditioned matrix-free LIE is equivalent to a linearly-implicit ROW method. The choice of the preconditioner, besides accelerating convergence, improves the stability of the matrix-free LIE method. The preconditioner defines the Jacobian approximation in the ROW method.*

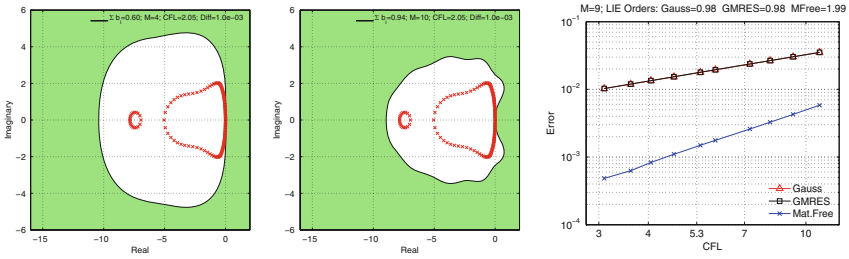
Note that the general approach can be applied to ROW methods [2, Sect. IV.7] by solving the linear system of each stage with an iterative matrix free algorithm. The resulting scheme is an explicit Runge Kutta method (or a ROW method) with $\sum_{i=1}^{s} m_i$ stages.

4 Numerical Results

Consider the one dimensional scalar advection-diffusion equation

$$u_t + (a\,u)_x = D\,u_{xx},\; u(x,t=0) = u_0(x)\,. \tag{10}$$

A spectral discontinuous Galerkin spatial discretization is used with 20 elements and polynomials of order 8. The diffusive term discretization is stabilized using the internal penalty method [8]. The LIE time stepping is used with the matrix-free GMRES solver [7].



(a) D=0.001, m=4 (b) D=0.001, m=10 (c) D=0.001, m=9

Fig. 1. (**a**) and (**b**) The ERK stability regions for different numbers of GMRES iterations. (**c**) The accuracy of the LIE scheme using various approaches to invert the Jacobian matrix. The GMRES weights are restricted by (7b) such as to obtain a second order method. Advection-diffusion equation (10), $\Delta t = CFL$, $\varepsilon = 10^{-6}\,\Delta t$

In Fig. 1a, b, the stability regions generated by the GMRES iterations are plotted for a varying number of Krylov vectors. The regions grow quickly and encompass the eigenvalues of the discrete advection-diffusion operator. Subsequent iterations improve solution accuracy but do not improve linear stability. Additional experiments (not reported here due to space constraints) reveal that the stability region of the resulting ERK method adapts to the eigenvalues of different discrete operators.

To verify the analysis in (7), we consider three different ways of computing the inverse of the linear Jacobian. The first is by Gauss elimination (LU), the second uses GMRES with the full Jacobian, and the third employs matrix-free GMRES iterations. In the last approach the GMRES coefficients are restricted by (7b) such as to obtain a second order time discretization method. Figure 1c shows the work-precision diagram for these approaches. The Gaussian elimination and traditional GMRES solutions display first order converge, while the constrained GMRES solution displays second order convergence.

5 Conclusions

Implicit time integration methods are becoming widely used in the the simulation of time dependent PDEs, as they do not suffer from CFL stability restrictions. While

implicit methods can use much larger time steps than explicit methods, their computational cost per step is also higher. The computational time is dominated by the solutions of (non)linear systems of equations that define each stage of a (linearly) implicit method. The implicit code is more effective only when the gains in step size offset the extra cost.

To reduce the computational overhead of LU decomposition, to alleviate memory requirements, and to aid parallelization, iterative Krylov space methods are used to solve the large linear systems. A matrix-free implementation approximates the required Jacobian vector products by finite differences.

This paper studies the effect of the matrix-free iterative solutions on the properties of the numerical integration method. The analysis reveals that matrix-free linearly implicit methods can be viewed as explicit Runge Kutta methods. Their stability region is finite, and the unconditional stability property of the original implicit method is lost. The equivalent Runge Kutta method is nonlinear, in the sense that its weights depend on the time step and on the stage vectors. This makes the accuracy analysis difficult. Order conditions of the equivalent explicit Runge Kutta method can be fulfilled by imposing additional conditions on the Krylov solution coefficients. For preconditioned matrix-free iterations the overall time stepping process is equivalent to a Rosenbrock-W method, where the preconditioner determines the Jacobian approximation. Future work will address the effect of a finite number of Krylov iterations on the stability and accuracy of the overall scheme, in the case where an analytical Jacobian is used.

Acknowledgments This work was supported by the NSF projects DMS–0915047, CCF–0916493, OCI–0904397, PetaApps–0904599. A. Sandu was also supported by NCAR's ASP Faculty Fellowship Program.

Bibliography

[1] A. Crivellini and F. Bassi. An implicit matrix-free Discontinuous Galerkin solver for viscous and turbulent aerodynamic simulations. *Computers and Fluids*, 50 (1):81–93, 2011.

[2] E. Hairer and G. Wanner. *Solving Ordinary Differential Equations II. Stiff and Differential-Algebraic Problems*. 2Ed. Springer-Verlag, 2002.

[3] E. Hairer, S.P. Norsett, and G. Wanner. *Solving Ordinary Differential Equations I. Nonstiff Problems.* Springer-Verlag, Berlin, 1993.

[4] Tim Kelley. *Iterative Methods for Optimization*. SIAM, 1999.

[5] D.A. Knoll and D.E. Keyes. Jacobian-free Newton-Krylov methods: a survey of approaches and applications. *Journal of Computational Physics*, 193(2):357–397, 2004.

[6] H. Luo, J.D. Baum, and R. Lohner. A fast, matrix-free implicit method for compressible flows on unstructured grids. In *Sixteenth international conference on numerical methods in fluid dynamics*, volume 515 of *Lecture Notes in Physics*, pages 73–78, 1998.

[7] Y. Saad. *Iterative Methods for Sparse Linear Systems*. SIAM, 2 edition, 2003.
[8] K. Shahbazi. An explicit expression for the penalty parameter of the interior penalty method. *Journal of Computational Physics*, 205:401–407, 2005.

Augmented Interface Systems for the Darcy-Stokes Problem

Marco Discacciati

Laboratori de Càlcul Numèric (LaCàN), Universitat Politècnica de Catalunya (UPC BarcelonaTech), Campus Nord UPC - C2, E-08034 Barcelona, Spain.
marco.discacciati@upc.edu

Summary. In this paper we study interface equations associated to the Darcy-Stokes problem using the classical Steklov-Poincaré approach and a new one called augmented. We compare these two families of methods and characterize at the discrete level suitable preconditioners with additive and multiplicative structures. Finally, we present some numerical results to assess their behavior in presence of small physical parameters.

1 Introduction and Problem Setting

Let $\Omega \subset \mathbb{R}^d$ $(d = 2,3)$ be a bounded domain decomposed into two non intersecting subdomains: Ω_f, filled by a viscous incompressible fluid, and Ω_p, formed by a porous medium, separated by an interface $\Gamma = \bar{\Omega}_f \cap \bar{\Omega}_p$. The fluid in Ω_f has no free surface and it can filtrate through the adjacent porous medium. The motion of the fluid in Ω_f is described by the Stokes equations:

$$-\nu\triangle\mathbf{u} + \nabla p = \mathbf{f}, \quad \operatorname{div}\mathbf{u} = 0 \quad \text{in } \Omega_f \tag{1}$$

where $\nu > 0$ is the kinematic viscosity, while $\mathbf{u}$ and p are the velocity and pressure. In Ω_p we describe the fluid motion by the equations:

$$\mathbf{u}_p = -\mathsf{K}\nabla\varphi, \quad \operatorname{div}\mathbf{u}_p = 0 \quad \text{in } \Omega_p \tag{2}$$

where $\mathbf{u}_p$ is the fluid velocity, φ the piezometric head and K the hydraulic conductivity tensor. The first equation is Darcy's law that provides the simplest linear relation between velocity and pressure in porous media. We can equivalently rewrite (2) as the elliptic equation involving only the piezometric head:

$$-\operatorname{div}(\mathsf{K}\nabla\varphi) = 0 \quad \text{in } \Omega_p. \tag{3}$$

Besides suitable boundary conditions on $\partial\Omega$, we supplement the Darcy-Stokes problem (1), (3) with the following coupling conditions on Γ:

$$-\mathsf{K}\nabla\varphi\cdot\mathbf{n} = \mathbf{u}\cdot\mathbf{n}, \quad -\mathbf{n}\cdot\mathsf{T}(\mathbf{u},p)\cdot\mathbf{n} = g\varphi, \quad -\varepsilon\boldsymbol{\tau}\cdot\mathsf{T}(\mathbf{u},p)\cdot\mathbf{n} = \nu\mathbf{u}\cdot\boldsymbol{\tau}, \tag{4}$$

R. Bank et al. (eds.), *Domain Decomposition Methods in Science and Engineering XX*, Lecture Notes in Computational Science and Engineering 91,
DOI 10.1007/978-3-642-35275-1_48,

where $\mathsf{T}(\mathbf{u},p)$ is the fluid stress tensor, $\boldsymbol{\tau}$ denotes a set of linear independent unit tangential vectors to Γ and ε is a coefficient related to the characteristic length of the pores of the porous medium. Conditions $(4)_1$ and $(4)_2$ impose the continuity of the normal velocity and of the normal component of the normal stress on Γ. The so-called Beavers-Joseph-Saffman condition $(4)_3$ does not yield any coupling but provides a boundary condition for the Stokes problem since it involves only quantities in the domain Ω_f. For more details we refer to [9, 11, 12, 14].

2 Interface Equations Associated to the Darcy-Stokes Problem

In [7, 8], we showed that the coupled Darcy-Stokes problem can be reformulated in terms of the solution of equations defined only on the interface Γ involving suitable Steklov-Poincaré operators associated to the subproblems in Ω_f and Ω_p. We formally briefly review this approach referring to the cited works for more details.

If we select as interface variable $\lambda \in H_{00}^{1/2}(\Gamma)$ to represent the normal velocity across Γ: $\lambda = \mathbf{u}\cdot\mathbf{n} = -\mathsf{K}\nabla\varphi\cdot\mathbf{n}$ on Γ, we can express the solution of the Darcy-Stokes problem in terms of the solution of the interface equation: find $\lambda \in H_{00}^{1/2}(\Gamma)$ such that

$$\langle S_s\lambda,\mu\rangle + \langle S_d\lambda,\mu\rangle = \langle \chi_s,\mu\rangle + \langle \chi_d,\mu\rangle \qquad \forall\mu \in H_{00}^{1/2}(\Gamma). \tag{5}$$

Equation (5) imposes the continuity condition $(4)_2$. The linear continuous operators χ_s and χ_d depend on the data of the problem and $\langle\cdot,\cdot\rangle$ denotes the duality pairing between $H_{00}^{1/2}(\Gamma)$ and its dual $(H_{00}^{1/2}(\Gamma))'$. Concerning S_s and S_d, we remark that

- The operator $S_s : H_{00}^{1/2}(\Gamma) \to (H_{00}^{1/2}(\Gamma))'$ maps the space of normal velocities on Γ to the space of normal stresses on Γ through the solution of a Stokes problem in Ω_f with boundary condition $\mathbf{u}\cdot\mathbf{n} = \lambda$ on Γ.
- S_d maps the space of fluxes of φ on Γ to the space of traces of φ on Γ via the solution of a Darcy problem in Ω_p with the boundary condition $-\mathsf{K}\nabla\varphi\cdot\mathbf{n} = \lambda$ on Γ. The operator S_d should be a map between $H^{-1/2}(\Gamma)$ and $H^{1/2}(\Gamma)$, but in (5) we are applying it to $H_{00}^{1/2}(\Gamma)$, a space with a higher regularity than needed where we cannot guarantee the coercivity of the operator.

On the other hand, if we choose as interface unknown $\eta \in H^{1/2}(\Gamma)$ the trace of the piezometric head on Γ: $\eta = g\varphi_{|\Gamma} = -\mathbf{n}\cdot\mathsf{T}(\mathbf{u},p)\cdot\mathbf{n}$ on Γ, the Darcy-Stokes problem can be equivalently reformulated as find $\eta \in H^{1/2}(\Gamma)$:

$$\langle\!\langle S_f\eta,\mu\rangle\!\rangle + \langle\!\langle S_p\eta,\mu\rangle\!\rangle = \langle\!\langle \chi_f,\mu\rangle\!\rangle + \langle\!\langle \chi_p,\mu\rangle\!\rangle \qquad \forall\mu \in H^{1/2}(\Gamma), \tag{6}$$

where χ_f and χ_p are linear continuous operators depending on the data of the problem. Equation (6) imposes the coupling condition $(4)_1$. Here:

- The operator S_f maps the space of normal stresses on Γ to the space of normal velocities on Γ via the solution of a Stokes problem with the boundary condition $-\mathbf{n}\cdot\mathsf{T}(\mathbf{u},p)\cdot\mathbf{n} = \eta$ on Γ. This operator would naturally be defined from

$H^{-1/2}(\Gamma)$ to $H_{00}^{1/2}(\Gamma)$ so that in (6) we are applying it to functions with a higher regularity than needed.

- The operator $S_p : H^{1/2}(\Gamma) \to (H^{1/2}(\Gamma))'$ maps the space of traces of φ on Γ to the space of fluxes of φ on Γ by solving a Darcy problem in Ω_p with the Dirichlet boundary condition $g\varphi = \eta$ on Γ.

3 Augmented Interface Equations

The classical approach summarized in Sect. 2 leads to reformulate the Darcy-Stokes problem as interface equations depending on a single interface unknown: either λ, the normal velocity across Γ, or η, the piezometric head on Γ. We have remarked that the Steklov-Poincaré operators S_d and S_f are not acting on their natural functional spaces, but they are assigned functions with higher regularity than expected. This prevents us from guaranteeing their coerciveness (see [7]). In this section we present a different approach based on [3–6] consisting in writing the coupled Darcy-Stokes problem as a system of linear equations on Γ involving both variables λ and η.

3.1 The Augmented Dirichlet-Dirichlet Problem

To obtain the augmented Dirichlet-Dirichlet (aDD) formulation assume that $\lambda \in H_{00}^{1/2}(\Gamma)$ is equal to the normal velocity $\mathbf{u}\cdot\mathbf{n}$ on Γ, but not necessarily to the conormal derivative of φ on Γ. On the other hand, let $\eta \in H^{1/2}(\Gamma)$ be equal to the trace of φ on Γ but not to the normal component of the Cauchy stress of the Stokes problem on Γ. Then, to recover the solution of the original Darcy-Stokes problem we have to impose both the continuity of normal velocity and of normal stresses:

$$\begin{aligned} -\textstyle\int_\Gamma \mathbf{n}\cdot\mathsf{T}(\mathbf{u}(\lambda),p(\lambda))\cdot\mathbf{n}\,\mu = \int_\Gamma \eta\mu \qquad & \forall \mu \in H_{00}^{1/2}(\Gamma) \\ -\textstyle\int_\Gamma \mathsf{K}\nabla\varphi(\eta)\cdot\mathbf{n}\,\xi = \int_\Gamma \lambda\xi \qquad & \forall \xi \in H^{1/2}(\Gamma). \end{aligned}$$

Using the definition of the Steklov-Poincaré operators, we can rewrite these conditions as: find $(\lambda,\eta) \in H_{00}^{1/2}(\Gamma)\times H^{1/2}(\Gamma)$ such that

$$\begin{aligned} \langle S_s\lambda,\mu\rangle + \langle\eta,\mu\rangle = \langle\chi_s,\mu\rangle \qquad & \forall \mu \in H_{00}^{1/2}(\Gamma) \\ \langle\!\langle S_p\eta,\xi\rangle\!\rangle - \langle\!\langle\lambda,\xi\rangle\!\rangle = \langle\!\langle\chi_p,\xi\rangle\!\rangle \qquad & \forall \xi \in H^{1/2}(\Gamma), \end{aligned} \tag{7}$$

or, in operator form:

$$\begin{pmatrix} S_s & \mathscr{I} \\ -\mathscr{J} & S_p \end{pmatrix}\begin{pmatrix}\lambda\\ \eta\end{pmatrix} = \begin{pmatrix}\chi_s\\ \chi_p\end{pmatrix} \tag{8}$$

where $\mathscr{I} : H^{1/2}(\Gamma) \to (H_{00}^{1/2}(\Gamma))'$ and $\mathscr{J} : H_{00}^{1/2}(\Gamma) \to (H^{1/2}(\Gamma))'$ are linear continuous maps.

We call (8) *augmented Dirichlet-Dirichlet* (aDD) formulation because both functions λ and η play the role of Dirichlet boundary conditions for the Stokes and the Darcy subproblems, respectively. Notice that we are imposing the equalities (8) in the sense of dual spaces and that the operators S_s and S_p still act on their natural functional spaces.

3.2 The Augmented Neumann-Neumann Problem

We follow now a similar approach to Sect. 3.1, but we assume that $\lambda \in H^{-1/2}(\Gamma)$ is equal to the conormal derivative of the piezometric head $-\mathrm{K}\nabla\varphi \cdot \mathbf{n}$ on Γ and $\eta \in H^{-1/2}(\Gamma)$ is equal to the normal component of the fluid Cauchy stress on Γ. Then, to recover the solution of the original problem we impose the following equalities:

$$\begin{aligned} &\textstyle\int_\Gamma \mathbf{u}(\eta)\cdot\mathbf{n}\,\mu = \int_\Gamma \lambda\,\mu \quad && \forall \mu \in H^{-1/2}(\Gamma) \\ &\textstyle\int_\Gamma \varphi(\lambda)\,\xi = -\int_\Gamma \eta\,\xi \quad && \forall \xi \in H^{-1/2}(\Gamma). \end{aligned}$$

Using the definition of the Steklov-Poincaré operators, we can rewrite these conditions as: find $(\lambda,\eta) \in H^{-1/2}(\Gamma)\times H^{-1/2}(\Gamma)$ such that

$$\begin{aligned} &\langle S_f\eta,\mu\rangle_* - \langle\lambda,\mu\rangle_* = \langle\chi_f,\mu\rangle_* \quad && \forall \mu \in H^{-1/2}(\Gamma) \\ &\langle\!\langle S_d\lambda,\xi\rangle\!\rangle_* + \langle\!\langle\eta,\xi\rangle\!\rangle_* = \langle\!\langle\chi_d,\xi\rangle\!\rangle_* \quad && \forall \xi \in H^{-1/2}(\Gamma), \end{aligned} \tag{9}$$

corresponding to the operator form:

$$\begin{pmatrix} S_d & \mathscr{I}_* \\ -\mathscr{I}_* & S_f \end{pmatrix} \begin{pmatrix} \lambda \\ \eta \end{pmatrix} = \begin{pmatrix} \chi_d \\ \chi_f \end{pmatrix}. \tag{10}$$

Here $\mathscr{I}_* : H^{-1/2}(\Gamma) \to H^{1/2}(\Gamma)$ and $\mathscr{J}_* : H^{-1/2}(\Gamma) \to H^{1/2}_{00}(\Gamma)$ are linear continuous maps, while $\langle\cdot,\cdot\rangle_*$ and $\langle\!\langle\cdot,\cdot\rangle\!\rangle_*$ denote the corresponding pairing.

We call this formulation *augmented Neumann-Neumann* (aNN) because both functions λ and η play the role of Neumann boundary conditions for the Darcy and the Stokes subproblems, respectively.

The aNN formulation may be regarded as the "dual" of the aDD approach. Notice that the operators S_f and S_d are now acting on their natural spaces, differently form the classical setting of Sect. 2. The analysis of problems (8) and (10) can be carried out following the guidelines of [5].

4 Algebraic Formulation of the Interface Problems

We consider a finite element discretization of the coupled problem using conforming grids across the interface Γ. The discrete spaces for the Stokes problem satisfy the inf-sup condition. In this way we obtain the linear system:

$$\begin{pmatrix} F & D & 0 & 0 \\ D^T & A_{\Gamma\Gamma} & 0 & -M_\Gamma \\ 0 & 0 & C_{ii} & C_{i\Gamma} \\ 0 & M_\Gamma^T & C_{\Gamma i} & C_{\Gamma\Gamma} \end{pmatrix} \begin{pmatrix} \boldsymbol{u}_i \\ \boldsymbol{u}_\Gamma \\ \boldsymbol{\varphi}_i \\ \boldsymbol{\varphi}_\Gamma \end{pmatrix} = \begin{pmatrix} \boldsymbol{f}_{fi} \\ \boldsymbol{f}_{f\Gamma} \\ \boldsymbol{f}_{pi} \\ \boldsymbol{f}_{p\Gamma} \end{pmatrix} \tag{11}$$

where $\boldsymbol{u}_\Gamma$ is the vector of the nodal values of the normal velocity on Γ while $\boldsymbol{u}_i$ is the vector of the remaining degrees of freedom (velocity and pressure) in Ω_f. On the other hand, $\boldsymbol{\varphi}_\Gamma$ is the vector of the (unknown) values of φ on Γ while $\boldsymbol{\varphi}_i$ corresponds to the remaining degrees of freedom in Ω_p.

The discrete counterpart of the Steklov-Poincaré operators can be found computing the Schur complement systems corresponding to either $\boldsymbol{u}_\Gamma$ or $\boldsymbol{\varphi}_\Gamma$. Precisely, we find:

$$\begin{aligned} \Sigma_s &= A_{\Gamma\Gamma} - D^T F^{-1} D, & \Sigma_f &= M_\Gamma^T \Sigma_s^{-1} M_\Gamma, \\ \Sigma_p &= C_{\Gamma\Gamma} - C_{\Gamma i} C_{ii}^{-1} C_{i\Gamma}, & \Sigma_d &= M_\Gamma \Sigma_p^{-1} M_\Gamma^T. \end{aligned} \tag{12}$$

The characterization of these discrete operators in terms of the associated Darcy or Stokes problems in Ω_p and Ω_f allows us to provide upper and lower bounds for their eigenvalues. Assuming ν and K constants in Ω_f and Ω_p, respectively, and the computational mesh to be uniform and regular, we can find (see [7, 13, 15]) ($\preceq$ indicates that the inequalities hold up to constants independent of h, ν, K):

$$\begin{aligned} h\nu &\preceq \sigma(\Sigma_s) \preceq \nu, & h^2\nu^{-1} &\preceq \sigma(\Sigma_f) \preceq h\nu^{-1} \\ h\mathsf{K} &\preceq \sigma(\Sigma_p) \preceq \mathsf{K}, & h^2\mathsf{K}^{-1} &\preceq \sigma(\Sigma_d) \preceq h\mathsf{K}^{-1} \end{aligned} \tag{13}$$

The discrete counterparts of the interface problems (5), (6), (8), and (10) read:

- Discrete interface equation for the normal velocity: find $\boldsymbol{u}_\Gamma$ such that

$$\Sigma_s \boldsymbol{u}_\Gamma + \Sigma_d \boldsymbol{u}_\Gamma = \boldsymbol{\chi}_s + \boldsymbol{\chi}_d. \tag{14}$$

- Discrete interface equation for the piezometric head: find $\boldsymbol{\varphi}_\Gamma$ such that

$$\Sigma_f \boldsymbol{\varphi}_\Gamma + \Sigma_p \boldsymbol{\varphi}_\Gamma = \boldsymbol{\chi}_f + \boldsymbol{\chi}_p. \tag{15}$$

- Discrete aDD problem: find $(\boldsymbol{u}_\Gamma, \boldsymbol{\varphi}_\Gamma)$ such that

$$\begin{pmatrix} \Sigma_s & -M_\Gamma \\ M_\Gamma^T & \Sigma_p \end{pmatrix} \begin{pmatrix} \boldsymbol{u}_\Gamma \\ \boldsymbol{\varphi}_\Gamma \end{pmatrix} = \begin{pmatrix} \boldsymbol{\chi}_s \\ \boldsymbol{\chi}_p \end{pmatrix}. \tag{16}$$

- Discrete aNN problem: find $(\boldsymbol{u}_\Gamma, \boldsymbol{\varphi}_\Gamma)$ such that

$$\begin{pmatrix} \Sigma_d & M_\Gamma \\ -M_\Gamma^T & \Sigma_f \end{pmatrix} \begin{pmatrix} \boldsymbol{u}_\Gamma \\ \boldsymbol{\varphi}_\Gamma \end{pmatrix} = \begin{pmatrix} \boldsymbol{\chi}_d \\ \boldsymbol{\chi}_f \end{pmatrix}. \tag{17}$$

The augmented approach allows to compute both interface variable at once but it requires to solve a system whose dimension is twice the one of the classical methods.

5 Iterative Solution Methods and Numerical Results

We present now some numerical methods to solve problems (14)–(17) focusing on cases where the fluid viscosity ν and the hydraulic conductivity K are small. These are indeed situations of interest for most practical applications. In [10] a Robin-Robin method was proposed to solve effectively (14). Here we adopt the generalized Hermitian/skew-Hermitian splitting (GHSS) method of [2] for (14) and (15) and the HSS method of [1] for (16) and (17). We start considering (14).

The matrix $\Sigma_s + \Sigma_d$ has no skew-symmetric component being symmetric positive definite, but thanks to the estimates (13) we can mimick the splitting proposed in [2]

considering Σ_s as a matrix multiplied by a coefficient (ν) which may become small. Thus, we can characterize the preconditioner for (14):

$$P_1 = (2\alpha_1)^{-1}(\Sigma_s + \alpha_1 I)(\Sigma_d + \alpha_1 I). \tag{18}$$

Proceeding analogously for (15), we can characterize the preconditioner

$$P_2 = (2\alpha_2)^{-1}(\Sigma_p + \alpha_2 I)(\Sigma_f + \alpha_2 I). \tag{19}$$

Preconditioners P_1 and P_2 involve suitable acceleration parameters α_1 and α_2 and can be used within GMRES iterations. Remark that they can be regarded as generalizations of the Robin-Robin method introduced in [7, 10].

On the other hand, as the matrices in (16) and (17) are positive skew-symmetric with symmetric positive definite diagonal blocks, we apply the HSS splitting proposed in [1] separating the symmetric and the skew-symmetric parts of the matrices. Thus, we can characterize the following preconditioners for GMRES iterations for (16) and (17), respectively, with α_3, α_4 suitable acceleration parameters:

$$P_3 = (2\alpha_3)^{-1} \begin{pmatrix} \Sigma_s + \alpha_3 I & 0 \\ 0 & \Sigma_p + \alpha_3 I \end{pmatrix} \begin{pmatrix} \alpha_3 I & -M_\Gamma \\ M_\Gamma^T & \alpha_3 I \end{pmatrix} \tag{20}$$

$$P_4 = (2\alpha_4)^{-1} \begin{pmatrix} \Sigma_d + \alpha_4 I & 0 \\ 0 & \Sigma_f + \alpha_4 I \end{pmatrix} \begin{pmatrix} \alpha_4 I & M_\Gamma \\ -M_\Gamma^T & \alpha_4 I \end{pmatrix}. \tag{21}$$

According to [2] these preconditioners are effective when either the skew-symmetric or the symmetric part dominates. Thanks to (13) we can expect that for small ν and K the skew-symmetric part dominates in (16) and the symmetric one in (17).

All preconditioners P_i require the solution of a Stokes problem in Ω_f and of a Darcy problem in Ω_p. However, P_1 and P_2 have a multiplicative structure while in P_3 and P_4 the two subproblems may be solved in a parallel fashion. They are all effective when ν and K become small. A thorough study of these preconditioners will make the object of a future work, where also the choice of the parameters α_i will be analyzed. For the tests reported in Table 1, following [2], we set $\alpha_1, \alpha_3 \simeq \sqrt{\nu}$, $\alpha_2 \simeq \sqrt{\mathsf{K}}$ and $\alpha_4 \simeq 10^{-1}$. However, a better characterization of such parameters is necessary to have a more robust behavior of the preconditioners, independent of both the mesh size and of the coefficients ν and K.

In the numerical tests, both the Stokes and the Darcy subproblems are solved via direct methods. The matrices in (20) and (21) involving M_Γ and I are assembled explicitly and the associated linear systems are solved using direct methods. We consider $\Omega_f = (0,1)\times(1,2)$, $\Omega_p = (0,1)^2$ with interface $\Gamma = (0,1)\times\{1\}$ and the analytic solution: $\mathbf{u} = ((y-1)^2 + (y-1) + 1, x(x-1))$, $p = 2\nu(x+y-1)$, $\varphi = \mathsf{K}^{-1}(x(1-x)(y-1) + (y-1)^3/3) + 2\nu x$. A comparison with preconditioners Σ_s for (14) and Σ_p for (15) studied in [7] is also presented. Although such preconditioners are optimal with unitary ν and K, they perform quite poorly when small viscosities and permeabilities are considered.

Bibliography

[1] Z.-Z. Bai, G.H. Golub, and M.K. Ng. Hermitian and skew-Hermitian splitting methods for non-Hermitian positive definite linear systems. *SIAM J. Matrix Anal. Appl.*, 24(3):603–626, 2003.

Table 1. Number of iterations to solve (14)-(17) using different preconditioners. Four computational meshes ($h_j = 2^{-(j+1)}$) and several values of ν and K have been considered.

GMRES iterations without and with preconditioner for (14) ($tol = 10^{-7}$).

	$\nu = 10^{-4}$, $\mathsf{K} = 10^{-3}$			$\nu = 10^{-6}$, $\mathsf{K} = 10^{-5}$			$\nu = 10^{-6}$, $\mathsf{K} = 10^{-8}$		
	No prec.	Σ_s	P_1	No prec.	Σ_s	P_1	No prec.	Σ_s	P_1
h_1	8	8	4 ($\alpha_1 = 10^{-2}$)	8	8	3 ($\alpha_1 = 10^{-3}$)	8	8	3 ($\alpha_1 = 10^{-3}$)
h_2	16	15	5 ($\alpha_1 = 10^{-2}$)	16	15	3 ($\alpha_1 = 10^{-3}$)	16	15	3 ($\alpha_1 = 10^{-3}$)
h_3	26	20	7 ($\alpha_1 = 10^{-3}$)	26	20	3 ($\alpha_1 = 10^{-3}$)	26	20	3 ($\alpha_1 = 10^{-3}$)
h_4	33	17	7 ($\alpha_1 = 10^{-3}$)	33	17	4 ($\alpha_1 = 10^{-3}$)	33	17	3 ($\alpha_1 = 10^{-3}$)

GMRES iterations without and with preconditioner for (15) ($tol = 10^{-7}$).

	$\nu = 10^{-4}$, $\mathsf{K} = 10^{-3}$			$\nu = 10^{-6}$, $\mathsf{K} = 10^{-5}$			$\nu = 10^{-6}$, $\mathsf{K} = 10^{-8}$		
	No prec.	Σ_p	P_2	No prec.	Σ_p	P_2	No prec.	Σ_p	P_2
h_1	9	9	6 ($\alpha_2 = 10^{-2}$)	9	9	4 ($\alpha_2 = 10^{-3}$)	-	-	3 ($\alpha_2 = 10^{-3}$)
h_2	17	17	7 ($\alpha_2 = 10^{-2}$)	17	17	4 ($\alpha_2 = 10^{-3}$)	-	-	3 ($\alpha_2 = 10^{-3}$)
h_3	32	31	8 ($\alpha_2 = 10^{-2}$)	33	33	5 ($\alpha_2 = 10^{-3}$)	33	33	4 ($\alpha_2 = 10^{-3}$)
h_4	46	42	8 ($\alpha_2 = 10^{-2}$)	59	57	5 ($\alpha_2 = 10^{-3}$)	63	62	4 ($\alpha_2 = 10^{-3}$)

GMRES iterations without and with preconditioner P_3 for (16) ($tol = 10^{-9}$).

	$\nu = 10^{-4}$, $\mathsf{K} = 10^{-3}$		$\nu = 10^{-6}$, $\mathsf{K} = 10^{-5}$		$\nu = 10^{-6}$, $\mathsf{K} = 10^{-8}$	
	No prec.	P_3	No prec.	P_3	No prec.	P_3
h_1	17	14 ($\alpha_3 = 10^{-3}$)	17	7 ($\alpha_3 = 10^{-3}$)	17	8 ($\alpha_3 = 10^{-3}$)
h_2	33	17 ($\alpha_3 = 10^{-3}$)	33	8 ($\alpha_3 = 10^{-3}$)	33	10 ($\alpha_3 = 10^{-3}$)
h_3	63	22 ($\alpha_3 = 5 \cdot 10^{-4}$)	65	8 ($\alpha_3 = 5 \cdot 10^{-4}$)	65	10 ($\alpha_3 = 5 \cdot 10^{-4}$)
h_4	67	23 ($\alpha_3 = 5 \cdot 10^{-4}$)	79	9 ($\alpha_3 = 5 \cdot 10^{-4}$)	101	11 ($\alpha_3 = 5 \cdot 10^{-4}$)

GMRES iterations without and with preconditioner P_4 for (17) ($tol = 10^{-9}$).

	$\nu = 10^{-4}$, $\mathsf{K} = 10^{-3}$		$\nu = 10^{-6}$, $\mathsf{K} = 10^{-5}$		$\nu = 10^{-6}$, $\mathsf{K} = 10^{-8}$	
	No prec.	P_4	No prec.	P_4	No prec.	P_4
h_1	17	16 ($\alpha_4 = 0.1$)	16	9 ($\alpha_4 = 0.5$)	9	8 ($\alpha_4 = 1$)
h_2	32	18 ($\alpha_4 = 0.1$)	32	8 ($\alpha_4 = 0.5$)	16	7 ($\alpha_4 = 0.5$)
h_3	59	20 ($\alpha_4 = 5 \cdot 10^{-2}$)	58	10 ($\alpha_4 = 0.1$)	30	5 ($\alpha_4 = 0.8$)
h_4	82	27 ($\alpha_4 = 5 \cdot 10^{-2}$)	81	8 ($\alpha_4 = 0.1$)	44	5 ($\alpha_4 = 0.8$)

[2] M. Benzi. A generalization of the Hermitian and skew-Hermitian splitting iteration. *SIAM J. Matrix Anal. Appl.*, 31(2):360–374, 2009.

[3] P.J. Blanco, R.A. Feijóo, and S.A. Urquiza. A unified variational approach for coupling 3D-1D models and its blood flow applications. *Comput. Methods Appl. Mech. Engrg.*, 196(41–44):4391–4410, 2007.

[4] P.J. Blanco, R.A. Feijóo, and S.A. Urquiza. A variational approach for coupling kinematically incompatible structural models. *Comput. Methods Appl. Mech. Engrg.*, 197(17–18):1577–1602, 2008.

[5] P.J. Blanco, M. Discacciati, and A. Quarteroni. Modeling dimensionally-heterogeneous problems: analysis, approximation and applications. *Numer. Math.*, 119(2):299–335, 2011.

[6] P.J. Blanco, P. Gervasio, and A. Quarteroni. Extended variational formulation for heterogeneous partial differential equations. *Computational Methods in Applied Mathematics*, 11(2):141–172, 2011.
[7] M. Discacciati. *Domain Decomposition Methods for the Coupling of Surface and Groundwater Flows*. PhD thesis, EPFL, 2004.
[8] M. Discacciati and A. Quarteroni. Analysis of a domain decomposition method for the coupling of Stokes and Darcy equations. In F. Brezzi et al., editor, *Numerical Mathematics and Advanced Applications, ENUMATH 2001*, pages 3–20. Springer. Milan, 2003.
[9] M. Discacciati and A. Quarteroni. Navier-Stokes/Darcy coupling: modeling, analysis, and numerical approximation. *Rev. Mat. Com.*, 22:315–426, 2009.
[10] M. Discacciati, A. Quarteroni, and A. Valli. Robin-Robin domain decomposition methods for the Stokes-Darcy coupling. *SIAM J. Numer. Anal.*, 45(3): 1246–1268, 2007.
[11] W. Jäger and A. Mikelić. On the boundary conditions at the contact interface between a porous medium and a free fluid. *Ann. Scuola Norm. Sup. Pisa Cl. Sci.*, 23:403–465, 1996.
[12] W. Jäger and A. Mikelić. On the interface boundary condition of Beavers, Joseph and Saffman. *SIAM J. Appl. Math.*, 60:1111–1127, 2000.
[13] L. Lakatos. Numerical analysis of iterative substructuring methods for the Stokes/Darcy problem. Master's thesis, EPFL, 2010.
[14] W.L. Layton, F. Schieweck, and I. Yotov. Coupling fluid flow with porous media flow. *SIAM J. Num. Anal.*, 40:2195–2218, 2003.
[15] A. Quarteroni and A. Valli. *Domain Decomposition Methods for Partial Differential Equations*. Oxford University Press, New York, 1999.

Mortar Coupling for Heterogeneous Partial Differential Equations

Pablo Blanco[1], Paola Gervasio[2], and Alfio Quarteroni[3]

[1] LNCC, Laboratório Nacional de Computação Científica, Av. Getúlio Vargas 333, Quitandinha, 25651-075 Petrópolis, RJ, Brazil. pjblanco@lncc.br
[2] DICATAM, University of Brescia, via Valotti, 9. 25133 Brescia, Italy. gervasio@ing.unibs.it
[3] MOX, Department of Mathematics "F. Brioschi", Politecnico di Milano, via Bonardi, 9. 20133 Milano, Italy and SB-MATHICSE-CMCS École Polytechnique Fédérale de Lausanne CH-1015 Lausanne, Switzerland. alfio.quarteroni@epfl.ch

1 Introduction

We are interested in the approximation of 2D elliptic equations with dominated advection and featuring boundary layers. In order to reduce the computational complexity, the domain is split into two subregions, the first one far from the layer, where we can neglect the viscosity effects, and the second one next to the layer. In the latter domain the original elliptic equation is solved, while in the former one, the pure convection equation obtained by the original one by dropping the diffusive term is approximated. The interface coupling is enforced by the non-conforming mortar method. We consider two different sets of interface conditions and we compare them for what concerns both computational efficiency and stability. One of the two sets of interface conditions turns out to be very effective, especially for very small viscosity when the mortar formulation of the original elliptic problem on the global domain can fail.

2 The Heterogeneous Problem

We consider an open bounded domain $\Omega \subset \mathbb{R}^2$ with Lipschitz boundary $\partial\Omega$, split into two open subsets Ω_1 and Ω_2 such that $\overline{\Omega} = \overline{\Omega}_1 \cup \overline{\Omega}_2, \Omega_1 \cap \Omega_2 = \emptyset$. Then, we denote by $\Gamma = \partial\Omega_1 \cap \partial\Omega_2$, the interface between the sub domains and we assume that Γ is of class $C^{1,1}$. Given $f \in L^2(\Omega)$, $b_0 \in L^\infty(\Omega)$, $\nu \in L^\infty(\Omega_2 \cup \Gamma)$ and $\mathbf{b} \in [W^{1,\infty}(\Omega)]^2$ satisfying the following inequalities:

$$\exists \nu_0 \in \mathbb{R} \text{ such that } \nu(\mathbf{x}) \geq \nu_0 > 0, \forall \mathbf{x} \in \Omega_2 \cup \Gamma,$$
$$\exists \sigma_0 \in \mathbb{R} \text{ such that } b_0(\mathbf{x}) + \tfrac{1}{2}\mathrm{div}\mathbf{b}(\mathbf{x}) \geq \sigma_0 > 0, \forall \mathbf{x} \in \Omega,$$

we look for two functions u_1 and u_2 (defined in $\overline{\Omega}_1$ and $\overline{\Omega}_2$, respectively) solutions of the *heterogeneous problem*

R. Bank et al. (eds.), *Domain Decomposition Methods in Science and Engineering XX*, Lecture Notes in Computational Science and Engineering 91,
DOI 10.1007/978-3-642-35275-1_49,

$$\begin{cases} \mathrm{div}(\mathbf{b}u_1) + b_0 u_1 = f & \text{in } \Omega_1, \\ \mathrm{div}(-\nu\nabla u_2 + \mathbf{b}u_2) + b_0 u_2 = f & \text{in } \Omega_2, \\ u_1 = 0 & \text{on } (\partial\Omega_1 \setminus \Gamma)^{in} \\ u_2 = 0 & \text{on } \partial\Omega_2 \setminus \Gamma \end{cases} \tag{1}$$

and satisfying the interface conditions

$$u_1 = u_2 \ \text{ on } \Gamma^{in}, \qquad \mathbf{b}\cdot\mathbf{n}_\Gamma u_1 + \nu\frac{\partial u_2}{\partial \mathbf{n}_\Gamma} - \mathbf{b}\cdot\mathbf{n}_\Gamma u_2 = 0, \ \text{ on } \Gamma. \tag{2}$$

$\mathbf{n}_\Gamma$ denotes the normal unit vector to Γ oriented from Ω_1 to Ω_2, while for any non-empty subset $S \subseteq \partial\Omega_1$, $S^{in} = \{\mathbf{x} \in S : \mathbf{b}(\mathbf{x})\cdot\mathbf{n}_1(\mathbf{x}) < 0\}$ and $S^{out} = \{\mathbf{x} \in S : \mathbf{b}(\mathbf{x})\cdot\mathbf{n}_1(\mathbf{x}) \geq 0\}$ are the *inflow* and the *outflow* parts of S, respectively.

Equations (2) (named IC1) express the continuity of the velocity field across the *inflow* part of the interface and the continuity of the fluxes across the whole interface. They can be equivalently expressed as (named IC2):

$$u_1 = u_2, \ \nu\frac{\partial u_2}{\partial \mathbf{n}_\Gamma} = 0 \ \text{ on } \Gamma^{in}, \quad -\mathbf{b}\cdot\mathbf{n}_\Gamma u_1 = \nu\frac{\partial u_2}{\partial \mathbf{n}_\Gamma} - \mathbf{b}\cdot\mathbf{n}_\Gamma u_2 \ \text{ on } \Gamma^{out}. \tag{3}$$

Problem (1) with either interface conditions (2) or (3) is well-posed, see [5].

The heterogeneous problem (1), with either interface conditions IC1 or IC2, can formally be written as an interface problem by means of Steklov-Poincaré operators (see, e.g., [3, 5]). Let us define the trace spaces $\Lambda_1 = L^2_{\mathbf{b}}(\Gamma^{in}) = \{v : \Gamma^{in} \to \mathbb{R} : \sqrt{|\mathbf{b}\cdot\mathbf{n}_\Gamma|}v \in L^2(\Gamma^{in})\}$ and $\Lambda_2 = H^{1/2}_{00}(\Gamma^{in}) = \{v : L^2(\Gamma^{in}) : \ \exists \tilde{v} \in H^{1/2}(\partial\Omega_2) : \tilde{v}|_{\Gamma^{in}} = v, \ \tilde{v}|_{\partial\Omega_2\setminus\Gamma^{in}} = 0\}$.

Solving (1) - (2) is equivalent to seeking $\lambda_k \in \Lambda_k$ for $k = 1, 2$, such that

$$\begin{cases} \mathscr{S}_1\lambda_1 + \mathscr{S}_2\lambda_2 = \chi_1 + \chi_2 & \text{in } \Lambda_2', \\ \lambda_1 = \lambda_2|_{\Gamma^{in}} & \text{in } \Lambda_2, \end{cases} \tag{4}$$

where

$$\mathscr{S}_1\lambda_1 = -\mathbf{b}\cdot\mathbf{n}_1 u_1^{\lambda_1}, \ \ \mathscr{S}_2\lambda_2 = \nu\frac{\partial u_2^{\lambda_2}}{\partial \mathbf{n}_2} - \mathbf{b}\cdot\mathbf{n}_2 u_2^{\lambda_2}, \quad \text{on } \Gamma, \tag{5}$$

are the local Steklov-Poincaré operators, while $u_1^{\lambda_1}$ and u_2^{λ} are the solution of

$$\begin{cases} \mathrm{div}(\mathbf{b}u_1^{\lambda_1}) + b_0 u_1^{\lambda_1} = 0 \ \text{ in } \Omega_1, \\ u_1^{\lambda_1} = 0 \text{ on } (\partial\Omega_1 \setminus \Gamma)^{in}, \quad u_1^{\lambda_1} = \lambda \text{ on } \Gamma^{in}, \end{cases} \tag{6}$$

and

$$\begin{cases} \mathrm{div}(-\nu\nabla u_2^{\lambda_2} + \mathbf{b}u_2^{\lambda_2}) + b_0 u_2^{\lambda_2} = 0 \text{ in } \Omega_2 \\ u_2^{\lambda_2} = 0 \text{ on } \partial\Omega_2 \setminus \Gamma, \quad u_2^{\lambda_2} = \lambda_2 \text{ on } \Gamma, \end{cases} \tag{7}$$

respectively. Finally,

$$\chi_1 = \mathbf{b}\cdot\mathbf{n}_1 u_1^f, \quad \chi_2 = -\nu\frac{\partial u_2^f}{\partial \mathbf{n}_2} + \mathbf{b}\cdot\mathbf{n}_2 u_2^f = -\nu\frac{\partial u_2^f}{\partial \mathbf{n}_2}, \tag{8}$$

where u_1^f and u_2^f are the solutions of problems like (6) and (7), respectively, with null trace on the interface and external load f. Note that $\chi_1|_{\Gamma^{in}} = 0$.

If interface conditions IC2 are considered instead of IC1, the resulting Steklov-Poincaré equation reads: seek $\lambda_k \in \Lambda_k$, for $k = 1,2$ such that

$$\begin{cases} \mathscr{S}_1^0\lambda_1 + \mathscr{S}_2^0\lambda_2 = \chi_1 + \chi_2 & \text{in } \Lambda_2' \\ \lambda_1 = \lambda_2|_{\Gamma^{in}} & \text{in } \Lambda_2 \end{cases} \tag{9}$$

where

$$\mathscr{S}_1^0\lambda_1 = \begin{cases} 0 & \text{on } \Gamma^{in} \\ -\mathbf{b}\cdot\mathbf{n}_1 u_1^{\lambda_1} & \text{on } \Gamma^{out}, \end{cases} \quad \mathscr{S}_2^0\lambda_2 = \begin{cases} \nu\dfrac{\partial u_2^{\lambda_2}}{\partial \mathbf{n}_2} & \text{on } \Gamma^{in} \\ \nu\dfrac{\partial u_2^{\lambda_2}}{\partial \mathbf{n}_2} - \mathbf{b}\cdot\mathbf{n}_2 u_2^{\lambda_2} & \text{on } \Gamma^{out}. \end{cases} \tag{10}$$

Remark 1. It is straightforward to prove that the operator $\mathscr{S}_2^0$ is always coercive on Λ_2, whereas $\mathscr{S}_2$ is coercive only if smallness assumption on $\mathbf{b}$ is assumed. If, e.g.,

$$\|\mathbf{b}\|_{L^\infty(\Gamma)} \le \varepsilon_0, \text{ with } 0 \le \varepsilon_0 \le 2\min\{\nu_0, \sigma_0\}/C_*^2, \tag{11}$$

(where C_* is the constant of the trace inequality $\|v\|_{L^2(\partial\Omega_2)} \le C_*\|v\|_{H^1(\Omega_2)}$) is satisfied then $\mathscr{S}_2$ is coercive on Λ_2. For this reason, the solution of problem (4) may produce oscillations around Γ^{in} when advection dominates (i.e. the global Péclet number is large), as will be shown later in our numerical results.

3 Mortar Coupling for Spectral Element Discretization

The discretization of the differential equation within each sub domain is performed by the quadrilateral conforming Spectral Element Method (SEM). We refer to [4] for a detailed description of this method. For $k = 1,2$, let $\mathscr{T}_k = \{T_{k,m}\}_{m=1}^{M_k}$ be a partition of the computational domain $\Omega_k \subset \mathbb{R}^2$. The SEM finite dimensional space on $\overline{\Omega}_k$ is denoted by X_{k,δ_k} and it is the set of functions in $C^0(\overline{\Omega}_k)$ whose restriction to $T_{k,m}$ is a polynomial of degree N_k in each direction. δ_k is an abridged notation for "discrete", that accounts for the local geometric sizes $h_{k,m}$ of $T_{k,m}$ and the local polynomial degrees N_k along each direction. Both geometric and polynomial conformity is guaranteed inside $\overline{\Omega}_k$.

The finite dimensional spaces in which we look for the SEM solution of either (4) or (9) are: $\Lambda_{1,\delta_1} \subset \Lambda_1$ and $\Lambda_{2,\delta_2} \subset \Lambda_2$. Their elements are globally continuous functions on Γ^{in} and Γ, respectively, and local polynomials of degree N_k on each edge induced by the partition $\mathscr{T}_k$.

For $k = 1,2$, we denote by $\mathscr{N}_{k,\Gamma}$ the set of nodes of $\mathscr{T}_k \cap \Gamma$ whose cardinality is $N_{k,\Gamma}$. Similar notations are used for the nodes lying on either Γ^{in} or Γ^{out}.

The finite dimensional basis $\{\mu_1^{(i)}\}_{i=1}^{N_{1,\Gamma^{in}}}$ of Λ_{1,δ_1} ($\{\mu_2^{(i)}\}_{i=1}^{N_{2,\Gamma}}$ of Λ_{2,δ_2}, resp.) is composed by the characteristic Lagrange polynomials in Ω_1 (Ω_2, resp.) associated to the Legendre-Gauss-Lobatto (LGL) nodes of $\mathcal{N}_{1,\Gamma^{in}}$ ($\mathcal{N}_{2,\Gamma}$, resp.). Then we set $(S_{2,\delta_2})_{ij} = \int_\Gamma \mathcal{S}_2 \mu_2^{(j)} \mu_2^{(i)} d\Gamma$ for $i,j = 0,\ldots,N_{2,\Gamma}$, and analogous notations are used to define matrices S^0_{2,δ_2}, S_{1,δ_1} and S^0_{1,δ_1}. Because of the high cost to compute integrals exactly, all integrals are approximated by Legendre-Gauss-Lobatto (LGL) quadrature rules.

We consider *non-conforming couplings*, i.e. we suppose that either the two partitions $\mathcal{T}_1$ and $\mathcal{T}_2$ do not share the same edges on Γ and/or the polynomial degrees do not coincide in the hyperbolic domain Ω_1 and in the elliptic one Ω_2. We adopt mortar methods (see, e.g., [2]) to glue non-conforming discretization across Γ.

The endpoints of the edges of $\mathcal{T}_1 \cap \Gamma^{in}$ are denoted by $v_1^{(i)}$, for $i = 1,\ldots,N_{1,v}$. $\tilde{\Lambda}_{1,\delta_1}$ is a suitable finite dimensional space of functions living on Γ^{in} and its basis functions ψ_l are characterized by being L^2 functions on Γ^{in} and local polynomials of degree $N_1 - 2$ on each edge of $\mathcal{T}_1 \cap \Gamma^{in}$. Therefore, the dimension of $\tilde{\Lambda}_{1,\delta_1}$ is $N_{\tilde{\Lambda}_1} = N_{1,\Gamma^{in}} - N_{1,v}$. By choosing Ω_2 as the master domain and Ω_1 as the slave, the continuity constraint $\lambda_1 = \lambda_2|_{\Gamma^{in}}$ is imposed weakly, i.e. by requiring that

$$\int_{\Gamma^{in}} (\lambda_{1,\delta_1} - \lambda_{2,\delta_2}) \psi_l d\Gamma = 0 \qquad \forall \psi_l \in \tilde{\Lambda}_{1,\delta_1}, \tag{12}$$

jointly with the strong continuity at the nodes $v_1^{(i)}$ of $\mathcal{T}_1 \cap \Gamma^{in}$, for $i = 1,\ldots,N_{1,v}$. This leads us to define a new set of *mortar* functions in Λ_{1,δ_1}, which are denoted by $\tilde{\mu}_1^{(k)}$ (for $k = 1,\ldots,N_{2,\Gamma^{in}}$) and satisfy the constraints:

$$\begin{cases} \tilde{\mu}_1^{(k)}(v_1^{(i)}) = \mu_2^{(k)}(v_1^{(i)}), & i = 1,\ldots,N_{1,v} \text{ and } v_1^{(i)} \text{ being endpoint} \\ & \text{of at least one edge of } \mathcal{T}_1 \cap \Gamma^{in} \\ \int_{\Gamma^{in}} (\tilde{\mu}_1^{(k)} - \mu_2^{(k)}) \psi_l d\Gamma = 0, & l = 1,\ldots,N_{\tilde{\Lambda}_1} \text{ and for all } \psi_l \in \tilde{\Lambda}_{1,\delta_1}. \end{cases} \tag{13}$$

Remark 2. We choose Ω_2 as the master domain because the nature of the heterogeneous problem requires to work with the trace of the elliptic solution on the whole interface and with the trace of the hyperbolic one only on Γ^{in}. Therefore it is more convenient to have the master trace at disposal on the whole Γ, instead of on a part of it.

The matrix form of system (13) reads

$$P\Xi = \Phi, \tag{14}$$

where $\Xi = [\xi_{jk}] \in \mathbb{R}^{N_{1,\Gamma^{in}} \times N_{2,\Gamma^{in}}}$ is defined by the relations

$$\tilde{\mu}_1^{(k)} = \sum_{j=1}^{N_{1,\Gamma^{in}}} \xi_{jk} \mu_1^{(j)}, \qquad k = 1,\ldots,N_{2,\Gamma^{in}}, \tag{15}$$

while $P \in \mathbb{R}^{N_{1,\Gamma^{in}} \times N_{1,\Gamma^{in}}}$ and $\Phi \in \mathbb{R}^{N_{1,\Gamma^{in}} \times N_{2,\Gamma^{in}}}$, are defined starting from (13). The matrix P is non-singular in view of the inf-sup condition for $\mathbb{Q}_N - \mathbb{Q}_{N-2}$ [2]. Once the discretization in Ω_1 and Ω_2 has been chosen, the matrix Ξ can be explicitly computed by solving (14).

The matrix Ξ enforces the gluing between degrees of freedom defined on $\mathcal{N}_{2,\Gamma^{in}}$ and $\mathcal{N}_{1,\Gamma^{in}}$. Therefore, Steklov-Poincaré equations (4) and (9) can be written in a nonconforming setting, by the use of matrix Ξ.

On Γ^{out} no continuity constraint, neither strong nor weak, is imposed, since the continuity of fluxes is a natural consequence of the interface equation. Nevertheless, on Γ^{out} we have to compute integrals of basis functions associated to two different meshes. To this aim we introduce the matrix $Q \in \mathbb{R}^{N_{2,\Gamma^{out}} \times N_{1,\Gamma^{out}}}$ for the evaluations of functions of Λ_{1,δ_1} at the nodes of $\mathcal{T}_2 \cap \Gamma$, and the matrix $D = M^{out}_{2,\delta_2} Q (M^{out}_{1,\delta_1})^{-1}$, where M^{out}_{k,δ_k} are the mass matrices induced by the LGL quadrature formulas on Γ^{out}, for $k = 1, 2$.

The nonconforming finite dimensional counterpart of (4) reads: find $\lambda_{k,\delta_k} \in \Lambda_{k,\delta_k}$ for $k = 1, 2$, such that

$$\begin{cases} \underbrace{\left(S_{2,\delta_2} + \begin{bmatrix} \Xi^T S^{in}_{1,\delta_1} \Xi & 0 \\ D S^{out}_{1,\delta_1} \Xi & 0 \end{bmatrix} \right)}_{S_\delta} \begin{bmatrix} \lambda^{in}_{2,\delta_2} \\ \lambda^{out}_{2,\delta_2} \end{bmatrix} = \begin{bmatrix} M^{in}_{2,\delta_2} \chi^{in}_{2,\delta_2} \\ M^{out}_{2,\delta_2} \chi^{out}_{2,\delta_2} + D \chi^{out}_{1,\delta_1} \end{bmatrix} \\ \lambda_{1,\delta} = \Xi \lambda^{in}_{2,\delta_2} \end{cases} \tag{16}$$

whereas that of (9) becomes: find $\lambda_{k,\delta_k} \in \Lambda_{k,\delta_k}$ for $k = 1, 2$, such that

$$\begin{cases} \underbrace{\left(S^0_{2,\delta_2} + \begin{bmatrix} 0 & 0 \\ D S^{out}_{1,\delta_1} \Xi & 0 \end{bmatrix} \right)}_{S^0_\delta} \begin{bmatrix} \lambda^{in}_{2,\delta_2} \\ \lambda^{out}_{2,\delta_2} \end{bmatrix} = \begin{bmatrix} M^{in}_{2,\delta_2} \chi^{in}_{2,\delta_2} \\ M^{out}_{2,\delta_2} \chi^{out}_{2,\delta_2} + D \chi^{out}_{1,\delta_1} \end{bmatrix} \\ \lambda_{1,\delta} = \Xi \lambda^{in}_{2,\delta_2}. \end{cases} \tag{17}$$

The upper scripts *in* and *out* denote the restriction to Γ^{in} and Γ^{out}, resp.

The numerical solutions of these linear systems is carried out by preconditioned Bi-CGStab iterations (see, [6]).

When conforming discretization is used across the interface (i.e. $\delta_1 = \delta_2$), matrix Ξ reduces to the identity matrix. In this situation, it is well known (see, e.g. [5]) that S^0_{2,δ_2} is an optimal preconditioner for the matrix S^0_δ, i.e. $\exists C_0 > 0$ independent of δ such that its spectral condition number $\mathcal{K}((S^0_{2,\delta_2})^{-1} S^0_\delta)$ is bounded by C_0. When $\delta_1 = \delta_2$, S^0_{2,δ_2} is an optimal preconditioner also for S_δ (see [3]), i.e. there exists $C_1 > 0$ independent of δ such that $\mathcal{K}((S^0_{2,\delta_2})^{-1} S_\delta) \le C_1$, and numerical results show that $C_0 \le C_1$.

We extend here the use of the preconditioner $S^0_{2,\delta}$ to the non-conforming case.

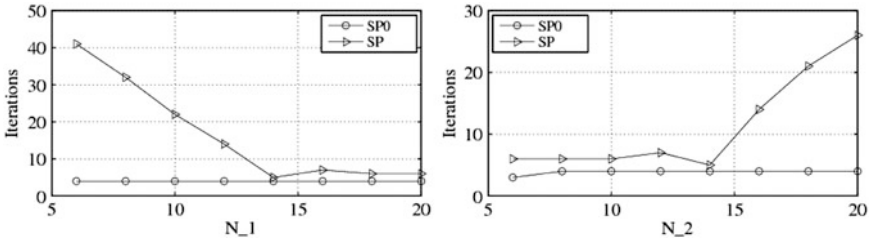


Fig. 1. Preconditioned Bi-CGStab iterations. The viscosity is $\nu = 10^{-2}$. *At left*, $N_2 = 14$ is fixed, *at right*, $N_1 = 14$ is fixed. 4×4 equal spectral elements are taken in each Ω_k

4 Numerical Results

Test case: the computational domain $\Omega = (-1,1)^2$ is split in $\Omega_1 = (-1,0.8) \times (-1,1)$ and $\Omega_2 = (0.8,1) \times (-1,1)$. The interface is $\Gamma = \{0.8\} \times (-1,1)$. The data of the problem are: $\mathbf{b} = [5y, 1-x]^t$, $b_0 = 1$, $f = 1$ and the inflow interface is $\Gamma^{in} = \{0.8\} \times (-1,0)$. The imposed Dirichlet boundary conditions are: $u_1 = 1$ on $((-1,0.8) \times \{-1\}) \cup (\{-1\} \times (0,1))$, $u_2 = 0$ on $\{1\} \times (-1,1)$, $u_2 = 1$ on $(0.8,1) \times \{-1\}$, while the homogeneous Neumann condition $\frac{\partial u_2}{\partial \mathbf{n}_2} = 0$ is imposed on $(0.8,1) \times \{1\}$.

Because of the presence of a boundary layer near the right vertical side, the mesh is refined there (without losing the conformity inside Ω_2) to prevent the numerical solution to be affected by spurious oscillations.

In Fig. 1 the number of Preconditioned Bi-CGStab (PBi-CGStab) iterations (with preconditioner S_{2,δ_2}) required to reduce the relative norm of the residual of 12 orders of magnitude is plotted versus the polynomial degrees N_1 and N_2 of the mortar discretization. These results refer to $\nu = 10^{-2}$ and show that the Steklov-Poincaré formulation (9) performs better than (4). The analysis of this and other test cases leads us to conjecture that $\mathcal{K}((S^0_{2,\delta_2})^{-1} S^0_\delta) \leq C_0$ still holds for non-conforming coupling ($\delta_1 \neq \delta_2$), while

$$\mathcal{K}((S^0_{2,\delta_2})^{-1} S_\delta) \simeq C_1 \mathcal{K}(\Xi \Xi^T) \simeq C_1 \begin{cases} (N_2 - N_1 + 1)^{3/2} & \text{if } N_1 < N_2 \\ C_2 & \text{if } N_1 \geq N_2, \end{cases} \tag{18}$$

where C_1 is the constant defined in the previous section, and C_2 is another positive constant independent of δ.

Therefore, formulation (17) corresponding to IC2 is optimally preconditioned by S^0_{2,δ_2} and it is better than (16) (corresponding to IC1) for what concerns the computational efficiency.

Moreover, when the viscosity vanishes (see Table 1), the performance of the SP0 approach (17) does not downgrade, as the number of PBi-CGStab iterations keeps bounded: three or four iterations are enough to satisfy the stopping test independently of both viscosity and discretization parameters.

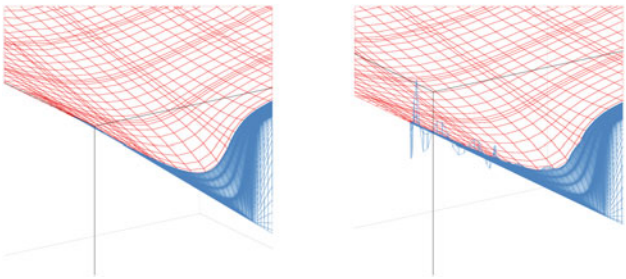


Fig. 2. Zoom on the numerical solution for $\nu = 10^{-3}$ and: (9) (*left*), (4) (*right*) with $N_1 = 8$ and $N_2 = 24$. The elliptic solution u_2 is in front, while the hyperbolic one u_1 is behind

On the contrary, the number of PBi-CGStab iterations required by SP approach (16) noticeably grows up when $\nu \to 0$ and behaves like $(N_2 - N_1 + 1)^{3/4}$ when $N_1 < N_2$, in agreement with (18).

The large number of PBi-CGStab iterations required by SP is due to the presence of instabilities across Γ^{in} which develop when advection dominates and the larger $N_2 - N_1$ is, the more they are pronounced.

We verified that the same instability occurs when mortar methods are applied to solve the pure elliptic-elliptic couplings with dominated advection and interface condition $\nu \frac{\partial u_1}{\partial \mathbf{n}_\Gamma} - \mathbf{b} \cdot \mathbf{n}_\Gamma u_1 = \nu \frac{\partial u_2}{\partial \mathbf{n}_\Gamma} - \mathbf{b} \cdot \mathbf{n}_\Gamma u_2$ on the whole interface Γ. Indeed, the local Steklov-Poincaré operators associated to the latter interface condition behaves like operator $\mathscr{S}_2$ introduced in (5), and they can lose the coercivity when $\|\mathbf{b}\|_{L^\infty(\Omega)}$ is large. This is the subject of a work in progress. (See also [1].)

In conclusion, the heterogeneous approach (1) with interface conditions IC2 and non-conforming mortar coupling turns out to be the most efficient and accurate one for vanishing viscosity and it is also a valid way to overcome instabilities arising from the mortar discretization of elliptic equations with dominated advection.

In Fig. 2 the heterogeneous solutions obtained by solving both (17) and (16) with $\nu = 10^{-4}$, $N_1 = 8$ and $N_2 = 24$ are shown. The elliptic solution u_2 provided by (16) (Fig. 2, right) exhibits non-trivial oscillations, while that provided by (17) (Fig. 2, left) does not.

Table 1. PBi-CGStab iterations to solve systems SP0 (17) and SP (16) with $P = S^0_{2,\delta_2}$ versus the viscosity. *At left,* $N_1 = 8$, *at right,* $N_1 = 20$, $N_2 = 24$. 4×4 equal spectral elements are taken in each Ω_k. $N_2 = 64$ along x-direction in the elements next to the layer

ν	10^{-1}	10^{-2}	10^{-3}	10^{-4}
SP0	3	4	3	3
SP	10	45	262	587

ν	10^{-1}	10^{-2}	10^{-3}	10^{-4}
SP0	3	3	3	4
SP	7	17	35	86

Bibliography

[1] Y. Achdou. The mortar element method for convection diffusion problems. *C.R. Acad. Sci. Paris. Sér. I Math.*, 321:117–123, 1995.

[2] C. Bernardi, Y. Maday, and A.T. Patera. A new nonconforming approach to domain decomposition: the mortar element method. In *Nonlinear partial differential equations and their applications. Collège de France Seminar, Vol. XI (Paris, 1989–1991)*, volume 299 of *Pitman Res. Notes Math. Ser.*, pages 13–51. Longman Sci. Tech., Harlow, 1994.

[3] P.J. Blanco, P. Gervasio, and A. Quarteroni. Extended variational formulation for heterogeneous partial differential equations. *Comp. Meth. in Applied Math.*, 11:141–172, 2011.

[4] C. Canuto, M. Y. Hussaini, A. Quarteroni, and T. A. Zang. *Spectral Methods. Evolution to Complex Geometries and Applications to Fluid Dynamics*. Springer, Heidelberg, 2007.

[5] F. Gastaldi, A. Quarteroni, and G. Sacchi Landriani. On the coupling of two dimensional hyperbolic and elliptic equations: analytical and numerical approach. In J.Périeaux T.F.Chan, R.Glowinski and O.B.Widlund, editors, *Third International Symposium on Domain Decomposition Methods for Partial Differential Equations*, pages 22–63, Philadelphia, 1990. SIAM.

[6] H.A. van der Vorst. *Iterative Krylov methods for large linear systems*, volume 13 of *Cambridge Monographs on Applied and Computational Mathematics*. Cambridge University Press, Cambridge, 2003.

Heterogeneous Substructuring Methods for Coupled Surface and Subsurface Flow

Heiko Berninger[1], Ralf Kornhuber[2], and Oliver Sander[2]

[1] Université de Genève, Section de Mathématiques, Switzerland,
Heiko.Berninger@unige.ch
[2] Freie Universität Berlin, Institut für Mathematik, Germany,
{kornhuber|sander}@math.fu-berlin.de

1 Introduction

The exchange of ground- and surface water plays a crucial role in a variety of practically relevant processes ranging from flood protection measures to preservation of ecosystem health in natural and human-impacted water resources systems.

Commonly accepted models are based on the shallow water equations for overland flow and the Richards equation for saturated–unsaturated subsurface flow with suitable coupling conditions. Continuity of mass flow across the interface is natural, because it directly follows from mass conservation. Continuity of pressure is typically imposed for simplicity. Mathematically, this makes sense for sufficiently smooth height of surface water as occurring, e.g., in filtration processes [9, 14]. Here we impose Robin-type coupling conditions modelling a thin, nearly impermeable layer at the bottom of the river bed that may cause pressure discontinuities; an effect which is known in hydrology as clogging (see [16] or [8, p. 1376]). From a mathematical perspective, clogging can be regarded as a kind of regularization, because, in contrast to Dirichlet conditions, Robin conditions can be straightforwardly formulated in a weak sense.

Existence and uniqueness results for the Richards equation and the shallow water equations are rare and hard to obtain, and nothing seems to be known about solvability of coupled problems. Extending the general framework of heterogeneous Steklov–Poincaré formulations and iterative substructuring [10, 13] to time-dependent problems, we introduce a Robin–Neumann iteration for the continuous coupled problem and motivate its feasibility by well-known existence results for the linear case. As surface and subsurface flow are only weakly coupled by clogging and continuity of mass flux, different discretizations with different time steps and different meshes can be used in a natural way. This is absolutely necessary, to resolve the vastly different time and length scales of surface and subsurface flow. Discrete mass conservation can be proved in a straightforward way.

Finally, we illustrate our considerations by coupling a finite element discretization of the Richards equation based on Kirchhoff transformation [4] with a simple

R. Bank et al. (eds.), *Domain Decomposition Methods in Science and Engineering XX*, 427
Lecture Notes in Computational Science and Engineering 91,
DOI 10.1007/978-3-642-35275-1_50,

upwind discretization of surface flow. Numerical experiments confirm discrete mass conservation and show fast convergence of the Robin–Neumann iteration for real-life soil data.

2 Coupled Surface and Subsurface Flow

Saturated–unsaturated subsurface flow during a time interval $(0, T_{\text{end}})$ in a porous medium occupying a bounded domain $\Omega \subset \mathbb{R}^d$, $d = 2, 3$, is described by the Richards equation

$$n\,\theta(p)_t + \operatorname{div} \mathbf{v}(p) = 0\,, \qquad \mathbf{v}(p) = -\frac{K}{\mu} kr(\theta(p))\nabla(p - \rho g z)\,. \tag{1}$$

The porosity n, permeability K, viscosity μ, and density ρ are given parameters, and g is the earth's gravitational acceleration. The unknown capillary pressure p is related to saturation $\theta(p)$ and relative permeability $kr(\theta(p))$ by equations of state [6, 7]

$$\theta(p) = \begin{cases} \theta_m + (\theta_M - \theta_m)\left(\frac{p}{p_b}\right)^{-\lambda} & \text{for } p \le p_b \\ \theta_M & \text{for } p \ge p_b \end{cases}$$

$$kr(\theta) = \left(\frac{\theta - \theta_m}{\theta_M - \theta_m}\right)^{3+\frac{2}{\lambda}}, \qquad \theta \in [\theta_m, \theta_M] \subset [0, 1]\,,$$

with residual saturation θ_m, maximal saturation θ_M, bubbling pressure $p_b < 0$, and pore size distribution factor $\lambda > 0$. Let $\Gamma \subset \partial\Omega$ denote the coupling boundary of the porous medium with a surface flow, and denote the outward normal vector of Γ by $\mathbf{n}$. We impose the coupling by Robin conditions $p|_\Gamma - \alpha \mathbf{v} \cdot \mathbf{n} \in L^2((0, T_{\text{end}}), H^{-1/2}(\Gamma))$ on Γ and homogeneous Neumann conditions on $\partial\Omega \setminus \Gamma$. With compatible initial conditions $\theta_0 \in L^1(\Omega)$ we assume that (1) admits a unique weak solution $p \in L^2((0, T_{\text{end}}), H^1(\Omega))$. This assumption is motivated by known existence results [1] for the Kirchhoff transformed Richards equation (see also [4]) and is, obviously, satisfied in the case of saturated flow $\theta \equiv \theta_M$.

The surface flow on Γ is described by the shallow water equations

$$h_t + \operatorname{div} \mathbf{q} = r, \tag{2a}$$

$$\mathbf{q}_t + \operatorname{div} \mathbf{F}(h, \mathbf{q}) = -gh\nabla\phi \tag{2b}$$

where $\phi : \Gamma_0 \to \Gamma$ is a parametrization of the surface topography of Γ. The unknown water height h and discharge $\mathbf{q}$, as well as a given mass source r are functions on $(0, T_{\text{end}}) \times \Gamma_0$. For ease of presentation, we assume $\Gamma = \Gamma_0$ so that Γ is an open subset of $\mathbb{R}^{d-1}$. For $d = 3$, i.e., $\Gamma \subset \mathbb{R}^2$, the flux function $\mathbf{F}$ takes the form

$$\mathbf{F} = \begin{pmatrix} \mathbf{F}_1 \\ \mathbf{F}_2 \end{pmatrix}, \qquad \mathbf{F}_1(h, \mathbf{q}) = \begin{pmatrix} q_1^2/h + \frac{1}{2}gh^2 \\ q_1 q_2 / h \end{pmatrix}, \quad \mathbf{F}_2(h, \mathbf{q}) = \begin{pmatrix} q_1 q_2 / h \\ q_2^2/h + \frac{1}{2}gh^2 \end{pmatrix}$$

with $\mathbf{q} = (q_1, q_2)$. It degenerates to $\mathbf{F}(h, \mathbf{q}) = \mathbf{q}^2/h + \frac{1}{2}gh^2$ for $\Gamma \subset \mathbb{R}$. For suitable initial conditions and inflow conditions on $\partial\Gamma_{\text{in}} \subset \partial\Gamma$ we assume that (2) has a weak solution $(h, \mathbf{q}) \in L^\infty((0, T_{\text{end}}), L^\infty(\Gamma))^d$ in the sense of distributions $\mathscr{D}'((0, T_{\text{end}}) \times \Gamma_{\text{in}})$ where $\Gamma_{\text{in}} = \Gamma \cup \partial\Gamma_{\text{in}}$. Since regularity results for nonlinear hyperbolic systems (2) do not seem to be available we note that this assumption is satisfied in the linear case [15, Theorem 2.2].

Mass conservation provides the Neumann coupling condition

$$r = \mathbf{v} \cdot \mathbf{n} .$$

Following, e.g. [16], we postulate a nearly impermeable river bed with thickness $\varepsilon \ll 1$ and permeability K_ε (clogging). Then Darcy's law provides the flux $\mathbf{v} = -\frac{K_\varepsilon}{\mu}\nabla p_\varepsilon$. Setting $\nabla p_\varepsilon = \varepsilon^{-1}(\rho g h - p|_\Gamma)\mathbf{n}$, we obtain the Robin coupling condition

$$p|_\Gamma - \alpha \mathbf{v} \cdot \mathbf{n} = \rho g h \tag{3}$$

with leakage coefficient $\alpha = \frac{\mu\varepsilon}{K_\varepsilon}$. Note that (3) generally implies a pressure discontinuity across the interface Γ between ground and surface water.

Remark 1. In light of the above regularity assumptions on pressure p and surface water height h coupling surface and subsurface flow by continuity $p|_\Gamma = \rho g h$ of capillary and hydrostatic pressure is generally not possible, because there is a regularity gap between the trace $p|_\Gamma \in L^2((0, T_{\text{end}}), H^{1/2}(\Gamma))$ and $h \in L^\infty((0, T_{\text{end}}), L^\infty(\Gamma)) \not\subset L^2((0, T_{\text{end}}), H^{1/2}(\Gamma))$ (see, e.g., [5, p. 148]) However, sufficient smoothness is available in special cases like, e.g., in- and exfiltration processes [14].

3 Steklov–Poincaré Formulation and Substructuring

We introduce the Robin-to-Neumann map

$$S_\Omega(h) = \mathbf{v}(h) \cdot \mathbf{n} = \alpha^{-1}(p|_\Gamma - \rho g h)$$

for $h \in L^\infty((0, T_{\text{end}}), L^\infty(\Gamma)) \subset L^2((0, T_{\text{end}}), H^{-1/2}(\Gamma))$. Here, p is the solution of the Richards equation (1) with Robin conditions (3). Assuming that for given $h \in L^\infty((0, T_{\text{end}}), L^\infty(\Gamma))$ and corresponding inflow boundary conditions, the second part (2b) of the shallow water equations has a unique weak solution $\mathbf{q}(h) \in L^\infty((0, T_{\text{end}}), L^\infty(\Gamma))^{d-1}$, we set

$$S_\Gamma(h) = -\operatorname{div} \mathbf{q}(h) .$$

The Steklov–Poincaré formulation of the coupled Richards equation and shallow water equations then reads

$$h_t = S_\Omega(h) + S_\Gamma(h) . \tag{4}$$

Just as (2a), the equality (4) is understood in the sense of distributions $\mathscr{D}'((0, T_{\text{end}}) \times \Gamma_{\text{in}})$.

In complete analogy to the stationary case [10, 13] we introduce a damped Robin–Neumann iteration

$$h_t^{\nu+1/2} - S_\Gamma(h^{\nu+1/2}) = S_\Omega(h^\nu)\,, \quad h^{\nu+1} = h^\nu + \omega(h^{\nu+1/2} - h^\nu)\,, \tag{5}$$

with a suitable damping parameter $\omega \in (0,\infty)$ and with an initial iterate given by $h^0 \in L^\infty((0,T_{\rm end}),L^\infty(\Gamma))$. Each step amounts to the solution of the Richards equation with Robin boundary conditions (3) to evaluate the source term $S_\Omega(h^\nu)$, and the subsequent solution of the shallow water equations (2) to evaluate $h^{\nu+1/2}$. The feasibility of (5) requires existence and uniqueness of these solutions. Note the similarity to waveform relaxation methods [11].

After selecting a step size $\Delta T = T_{\rm end}/N$ with suitable $N \in \mathbb{N}$ and corresponding time levels $T_k = k\Delta T$, the Robin–Neumann iteration (5) can also be applied on subintervals $[T_{k-1},T_k]$, $k = 1,\dots,N$.

4 Discretization and Discrete Robin–Neumann Iteration

We first derive a discrete version of the Steklov–Poincaré formulation (4) on a fixed time interval $[T_k,T_{k+1}]$ with $0 \le T_k < T_{k+1} = T_k + \Delta T \le T_{\rm end}$. To this end, we introduce intermediate time levels $t_i = T_k + i\tau$, $i = 0,\dots,M$, with step size $\tau = \Delta T/M$ and suitable $M \in \mathbb{N}$. Spatial discretization is based on a partition $\mathscr{T}_\Gamma$ of Γ into simplices T that is regular in the sense that the intersection of two simplices T, $T' \in \mathscr{T}_\Gamma$ is either a common face, edge, vertex, or empty. We introduce the corresponding space of discontinuous finite elements of order $q \ge 0$ by

$$\mathscr{V}_\Gamma = \{v \in L^2(\Gamma) \mid v_T \text{ is a polynomial of degree at most } q \;\forall T \in \mathscr{T}_\Gamma\}\,,$$

and let $h = (h_i)_{i=0}^M$ denote approximations $h_i \in \mathscr{V}_\Gamma$ at t_i, $i = 0,\dots,M$.

Then, utilizing the forward difference quotient $\partial_t h_i = (h_{i+1} - h_i)/\tau$, a discrete Steklov–Poincaré formulation reads

$$\partial_t h_i = S_\Gamma(h)_i + S_\Omega(h)_i, \qquad i = 0,\dots,M-1\,. \tag{6}$$

Here and in the rest of this section, subscripts i indicate approximations taken at time t_i.

For given $h = (h_i)_{i=0}^M$, the discrete surface flow

$$(S_\Gamma(h)_i, v)_\Gamma = \sum_{T\in\mathscr{T}_\Gamma} \big((\mathbf{q}(h)_i, \nabla v)_T + (\mathbf{G}_h(h_i,\mathbf{q}(h)_i)\cdot\mathbf{n}_T, v)_{\partial T}\big) \quad \forall v \in \mathscr{V}_\Gamma \tag{7}$$

results from an explicit discontinuous Galerkin discretization of (2a), characterized by the discrete flux function $\mathbf{G}_h$. Here, $(\cdot,\cdot)_U$ stands for the L^2 scalar product on $U = \Gamma$, T, ∂T, respectively; $\mathbf{n}_T$ is the outward normal on T, and the discrete discharge $\mathbf{q}_i = \mathbf{q}(h)_i$ is obtained from an explicit discontinuous Galerkin discretization of (2b)

$$(\partial_t \mathbf{q}_i, v)_\Gamma = \sum_{T\in\mathscr{T}_\Gamma} \big((\mathbf{F}(h_i,\mathbf{q}_i), \nabla v)_T + (\mathbf{G}_\mathbf{q}(h_i,\mathbf{q}_i)\cdot\mathbf{n}_T, v)_{\partial T}\big) \quad \forall v \in (\mathscr{V}_\Gamma)^{d-1}\,. \tag{8}$$

Since we expect the dynamics of subsurface flow to be much slower than the surface water dynamics, we use the macro time step ΔT for an implicit time discretization of $S_\Omega(h)$. The spatial discretization is based on conforming piecewise linear finite elements

$$\mathscr{V}_\Omega = \{v \in C(\overline{\Omega}) \mid v|_T \text{ is affine linear } \forall T \in \mathscr{T}_\Omega\}$$

with respect to a regular partition $\mathscr{T}_\Omega$ of Ω. No compatibility conditions on $\mathscr{T}_\Omega$ and $\mathscr{T}_\Gamma$ are required. For given $p_k \in \mathscr{V}_\Omega$ and $h_{k+1} \in \mathscr{V}_\Gamma$, the discrete capillary pressure $p_{k+1} \in \mathscr{V}_\Omega$ is then obtained from the variational equality

$$\begin{aligned} n\langle\theta_{k+1}, v\rangle_\Omega + \Delta T\big((\mathbf{v}_{k+1}, \nabla v)_\Omega \\ +\alpha^{-1}(\langle p_{k+1}|_\Gamma, v\rangle_\Gamma - (\rho g h_{k+1}, v)_\Gamma)\big) = n\langle\theta_k, v\rangle_\Omega \qquad \forall v \in \mathscr{V}_\Omega. \end{aligned} \tag{9}$$

Here $\langle\cdot,\cdot\rangle_\Omega$ denotes the lumped L^2 scalar product on Ω, $\langle\cdot,\cdot\rangle_\Gamma$ is the corresponding lumped L^2 scalar product on Γ, $\theta_k = \theta(p_k)$, and $\mathbf{v}_{k+1}$ is a discretization of the flux $\mathbf{v}$ at T_{k+1}. Once $p_{k+1} \in \mathscr{V}_\Omega$ is available, we set for all $i = 0, \ldots, M$

$$(S_\Omega(h)_i, v)_\Gamma = \alpha^{-1}(p_{k+1}|_\Gamma - \rho g h_{k+1}, v)_\Gamma \qquad \forall v \in \mathscr{V}_\Gamma\,. \tag{10}$$

Note that $S_\Omega(h)_i$ is constant on the macro interval $[T_k, T_{k+1}]$ and only depends on h_{k+1}.

Testing (6) and (9) with constant functions $\mathbf{1} \in \mathscr{V}_\Gamma$ and $\mathbf{1} \in \mathscr{V}_\Omega$, respectively, and using $\langle p_{k+1}|_\Gamma, \mathbf{1}\rangle_\Gamma = (p_{k+1}|_\Gamma, \mathbf{1})_\Gamma$ we obtain discrete mass conservation.

Proposition 1. *The discrete Steklov–Poincaré formulation* (6) *with* S_Γ *and* S_Ω *defined by* (7) *and* (10) *is mass conserving in the sense that*

$$(h_{k+1}, \mathbf{1})_\Gamma + n\langle\theta_{k+1}, \mathbf{1}\rangle_\Omega = (h_k, \mathbf{1})_\Gamma + n\langle\theta_k, \mathbf{1}\rangle_\Omega + \tau \sum_{i=0}^{M-1} (\mathbf{G}_h(h_i, \mathbf{q}_i) \cdot \mathbf{n}_{\partial\Gamma}, \mathbf{1})_{\partial\Gamma}$$

holds for $k = 0, 1, \ldots,$ *with* $\mathbf{n}_{\partial\Gamma}$ *denoting the outward normal on* $\partial\Gamma$.

We emphasize that this result holds for arbitrary discretizations of the Richards flux $\mathbf{v}$.

The discrete Steklov–Poincaré formulation (6) gives rise to the discrete damped Robin–Neumann iteration

$$\partial_t h_i^{\nu+1/2} - S_\Gamma(h^{\nu+1/2})_i = S_\Omega(h^\nu)_i\,, \quad h_i^{\nu+1} = h_i^\nu + \omega(h_i^{\nu+1/2} - h_i^\nu)\,, \tag{11}$$

with suitable damping parameter $\omega \in (0, \infty)$, and an initial iterate $h_i^0 \in \mathscr{V}_\Gamma$ for $i = 0, \ldots, M$. Each step amounts to the solution of the discretized Richards equation (9) to obtain $S_\Omega(h^\nu)_i$ from (10) with $p_{k+1} = p_{k+1}^{\nu+1}$, and to M time steps of the discontinuous Galerkin discretization of (2) described by (7) and (8) to obtain $h_i^{\nu+1/2}$, $i = 1, \ldots, M$. For $k > 0$ the initial iterate h^0 is the solution of the preceding time step. We emphasize that no compatibility conditions on the different meshes $\mathscr{T}_\Gamma$ and $\mathscr{T}_\Omega$ are necessary, because only weak coupling conditions are involved.

5 Numerical Experiments

We consider a model problem on a square $\Omega \subset \mathbb{R}^2$ of side length 10 m and select Γ as the upper part of its boundary. The soil parameters are $n = 0.437$, $\theta_m = 0.0458$, $\theta_M = 1$, $p_b = -712.2\,\mathrm{Pa}$, $\lambda = 0.694$, and $K = 6.66 \cdot 10^{-9}\,\mathrm{m}^2$ (sandy soil). The viscosity and density of water is $\mu = 1\,\mathrm{m\,Pa\,s}$ and $\rho = 1{,}000\,\mathrm{kg\,m}^{-3}$, respectively. In accordance with measurements [16] we select the leakage coefficient as $\alpha = \rho g L^{-1}$ with $L = 10^{-6}\,\mathrm{s}^{-1}$ allowing for large pressure jumps across the interface.

We choose the initial conditions $\theta_0 \equiv \theta(-20\,\mathrm{Pa}) = 0.1401$, $h(0) \equiv 1\,\mathrm{m}$, $\mathbf{q}(0) \equiv 10\,\mathrm{m}^2\,\mathrm{s}^{-1}$, and inflow boundary conditions for $h(0,t)$ and $\mathbf{q}(0,t)$ alternating between 2 and 1 m and 20 and $10\,\mathrm{m}^2\,\mathrm{s}^{-1}$, respectively, with a period of 10 s. This leads to a supercritical water flow from left to right, which can result, for example, from opening a flood gate.

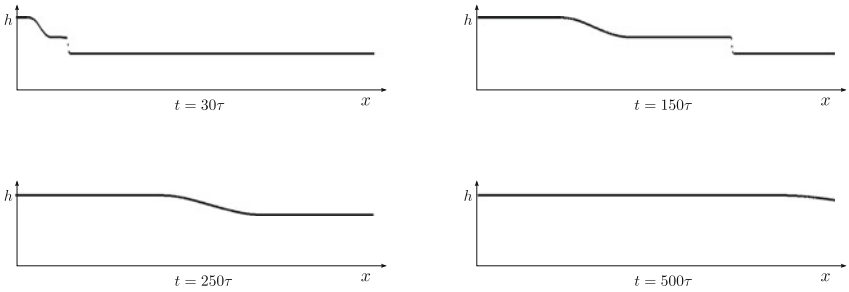


Fig. 1. The water height h_i at times $t_i = i\tau$, $i = 30, 150, 250, 500$

Fig. 2. The pressure p at times $T_k = k\Delta T$, $k = 200, 1000, 2000, 3000$

For the porous media flow on Ω we use the uniform time step size $\Delta T = 50\,\mathrm{s}$ and a triangulation $\mathcal{T}_\Omega$ resulting from six uniform refinement steps applied to a partition of Ω into two triangles with hypotenuse from lower left to upper right. The Richards equation (1) is discretized by the implicit scheme based on Kirchhoff transformation suggested in [4], and truncated monotone multigrid [12] is used as the algebraic solver. For the surface flow we use the time step size $\tau = \gamma\Delta T$ with $\gamma = 3^{-1} \cdot 10^{-4}$,

and the partition $\mathscr{T}_\Gamma$ consists of 400 elements of equal length. Note that $\mathscr{T}_\Gamma$ does not match with $\mathscr{T}_\Omega|_\Gamma$. The shallow water equations (2) are discretized by a discontinuous Galerkin method (7) with $\mathscr{V}_\Gamma$ consisting of piecewise constant functions, and we use simple upwind flux functions $\mathbf{G}_h$ and $\mathbf{G}_\mathbf{q}$ in (7) and (8), respectively. The final time is $T_{\text{end}} = 3.5 \cdot 10^4$ s. For the implementation we used the DUNE libraries [2] and the domain decomposition module `dune-grid-glue` [3].

Figure 1 shows the evolution of the surface water height h over the first period of the boundary conditions. The porous medium flow is much slower, as expected. Figure 2 shows the evolution of the pressure. Water enters the domain from the top, and after about 3,600 macro time steps or, equivalently, 3,000 m, the soil saturation is constant at about 75 %. Then, the domain gets fully saturated starting from the bottom. Hydrostatic pressure builds up and is fully reached at time step 4,700.

At each time step we observe discrete mass conservation up to machine precision. The total relative mass loss over the entire evolution is about 10^{-10}. Our numerical computations thus nicely reproduce the theoretical findings of Proposition 1.

In order to investigate the convergence behavior of the Robin–Neumann iteration (11), we consider the algebraic error $\|h_M - h_M^\nu\|_{L^1(\Gamma)}$ at the end of the first time interval $[0, T_1]$ with $T_1 = M\tau$. It turns out that for the given leakage coefficient $\alpha = \rho g 10^6$ s (cf. [16]), the convergence rates are in the range of 10^{-4}. They remain there during the entire evolution. For each time step only two or three iterations were necessary to reduce the estimated algebraic error below the threshold 10^{-12}. This is explained by the weak (in the physical sense) coupling of surface water and subsurface flow associated with large values of α.

The convergence speed of (11) decreases for decreasing α. This is illustrated in Fig. 3 which shows convergence rates ρ of (11) for various α together with the corresponding optimal damping factors ω determined numerically. Convergence rates

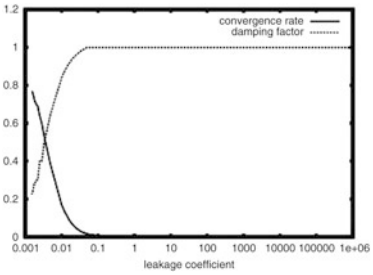


Fig. 3. Convergence rates ρ and associated optimal damping parameter ω over leakage coefficient α

deteriorate for $\alpha < 4 \cdot 10^{-2}$. Moreover, for $\alpha < 2 \cdot 10^{-3}$ ill-conditioning of the discretized Richards equation (9) leads to severe problems in the numerical solution. Hence, using the Robin coupling (3) to enforce continuity of pressure by penaliza-

tion rather than for modelling the clogging effect would require the construction of suitable preconditioners and a careful selection of α.

Bibliography

[1] H.W. Alt and S. Luckhaus. Quasilinear elliptic–parabolic differential equations. *Math. Z.*, 183:311–341, 1983.

[2] P. Bastian, M. Blatt, A. Dedner, C. Engwer, R. Klöfkorn, R. Kornhuber, M. Ohlberger, and O. Sander. A generic interface for parallel and adaptive scientific computing. Part II: Implementation and tests in DUNE. *Computing*, 82(2–3):121–138, 2008.

[3] P. Bastian, G. Buse, and O. Sander. Infrastructure for the coupling of Dune grids. In *Proc. of ENUMATH 2009*, pages 107–114. Springer, 2010.

[4] H. Berninger, R. Kornhuber, and O. Sander. Fast and robust numerical solution of the Richards equation in homogeneous soil. Technical Report A /01/2010, FU Berlin, 2010. submitted to SIAM J. Numer. Anal.

[5] F. Brezzi and G. Gilardi. Functional spaces. In H. Kardestuncer and D.H. Norrie, editors, *Finite Element Handbook*, chapter 2 (part 1), pages 1.29–1.75. Springer, 1987.

[6] R.J. Brooks and A.T. Corey. Hydraulic properties of porous media. Technical Report Hydrology Paper No. 3, Colorado State University, 1964.

[7] N.T. Burdine. Relative permeability calculations from pore-size distribution data. *Petr. Trans., Am. Inst. Mining Metall. Eng.*, 198:71–77, 1953.

[8] C. Dawson. Analysis of discontinuous finite element methods for ground water/surface water coupling. *SIAM J. Numer. Anal.*, 44(4):1375–1404, 2006.

[9] C. Dawson. A continuous/discontinuous Galerkin framework for modeling coupled subsurface and surface water flow. *Comput. Geosci*, 12:451–472, 2008.

[10] S. Deparis, M. Discacciati, G. Fourestey, and A. Quarteroni. Fluid–structure algorithms based on Steklov–Poincaré operators. *Comput. Methods Appl. Mech. Engrg.*, 195:5797–5812, 2008.

[11] L. Halpern. Schwarz waveform relaxation algorithms. In *Domain Decomposition Methods in Science and Engineering XVII*, volume 60 of *Lecture Notes in Computational Science and Engineering*, pages 155–164. Springer, 2008.

[12] R. Kornhuber. On constrained Newton linearization and multigrid for variational inequalities. *Numer. Math.*, 91:699–721, 2002.

[13] A. Quarteroni and A. Valli. *Domain Decomposition Methods for Partial Differential Equations*. Oxford Science Publications, 1999.

[14] P. Sochala, A. Ern, and S. Piperno. Mass conservative BDF-discontinuous Galerkin/explicit finite volume schemes for coupling subsurface and overland flows. *Comput. Methods Appl. Mech. Engrg.*, 198:2122–2136, 2009.

[15] N.J. Walkington. Convergence of the discontinuous Galerkin method for discontinuous solutions. *SIAM J. Numer. Anal.*, 42:1801–1817, 2005.

[16] B. Wiese and G. Nützmann. Transient leakance and infiltration characteristics during lake bank filtration. *Ground Water*, 47(1):57–68, 2009.

An Asymptotic Approach to Compare Coupling Mechanisms for Different Partial Differential Equations

Martin J. Gander[1] and Véronique Martin[2]

[1] Section de Mathématiques, Université de Genève, 2-4 rue du Lièvre, CP 64, CH-1211 Genève, SWITZERLAND. martin.gander@unige.ch
[2] LAMFA UMR 7352 Université de Picardie Jules Verne, Amiens, FRANCE. veronique.martin@u-picardie.fr

1 Introduction

In many applications the viscous terms become only important in parts of the computational domain. A typical example is the flow of air around the wing of an airplane. It can then be desirable to use an expensive viscous model only where the viscosity is essential for the solution and an inviscid one elsewhere. This leads to the interesting problem of coupling partial differential equations of different types.

The purpose of this paper is to explain several coupling strategies developed over the last decades, and to introduce a systematic way to compare them. We will use the following simple model problem to do so:

$$\begin{aligned} \mathscr{L}_{ad}u := -\nu u'' + au' + cu &= f \quad \text{in } \Omega = (-L_1, L_2), \\ \mathscr{B}_1 u &= g_1 \text{ on } x = -L_1, \\ \mathscr{B}_2 u &= g_2 \text{ on } x = L_2, \end{aligned} \tag{1}$$

where ν and c are strictly positive constants, $a, g_1, g_2 \in \mathbf{R}$, $f \in L^2(\Omega)$, $L_1, L_2 > 0$ and $\mathscr{B}_j$, $j = 1,2$ are suitable boundary operators of Dirichlet, Neumann or Robin type. If in part of Ω, the diffusion plays only a minor role, one would like to replace the viscous solution u by an inviscid approximation, which leads to two separate problems: a viscous problem on, say, $\Omega^- := (-L_1, x_0 + \delta)$, where δ stands for the size of the overlap and x_0 the position of the interface,

$$\begin{aligned} \mathscr{L}_{ad}u_{ad} &= f \quad \text{in } \Omega^-, \\ \mathscr{B}_1 u_{ad} &= g_1 \text{ on } x = -L_1, \end{aligned} \tag{2}$$

and a pure advection reaction problem on $\Omega^+ := (x_0, L_2)$,

$$\mathscr{L}_a u_a := a u_a' + c u_a = f \qquad \text{in } \Omega^+. \tag{3}$$

Coupling conditions for (2) and (3) need then to be chosen to connect the two subproblems, and there are many coupling strategies in the literature to choose from.

R. Bank et al. (eds.), *Domain Decomposition Methods in Science and Engineering XX*, Lecture Notes in Computational Science and Engineering 91,
DOI 10.1007/978-3-642-35275-1_51,

These strategies have been developed over the last decades for various applications, and sometimes the two different models are really due to different physical phenomena, like in fluid-structure interaction problems. In those cases, the coupling conditions are given by the physics, and they are in general unique. We are however interested in problems where the different equations are only chosen in order to achieve computational savings, as for example in [5]:

> The main goal of this paper is to present a computational method for the coupling of two distinct mathematical models describing the same physical phenomenon.

For such couplings, it is quite difficult to decide which coupling strategy from the literature to choose, since every coupling strategy leads to a different solution, and it is not clear a priori which one is the best one. Furthermore, there are neither guidelines nor quantitative comparisons in the literature in order to help with this decision. In order to compare the quality of the various coupling strategies, we propose in this paper a first very natural measure to compare different coupling strategies in such situations, namely to investigate how close the coupled solution for (2) and (3) is to the fully viscous solution of (1). The idea behind this quality measure is that in principle the viscosity should be taken into account everywhere, and hence it is the more expensive viscous solution that we are interested in. However, for computational savings, one would like to use a simpler, non-viscous model whenever the viscosity does not play an important role. In a more general situation, we thus would propose as a natural quality measure to compare the coupled solution to the solution of the expensive model used throughout the entire domain, and the closer the coupled solution is to this expensive one, the better the coupling conditions are.

We describe in this paper in detail several coupling strategies for the viscous/inviscid coupling, and compare them by testing how close the coupled solution is to the fully viscous one: in Sect. 2 we present an overlapping coupling method based on optimization. In Sect. 3 we present several non-overlapping coupling strategies based on coupling conditions at the interface between the two regions. In both sections, the position of the interface needs to be known a priori. This is in contrast to Sect. 4, where we present an adaptive coupling strategy which detects the partition into viscous and non-viscous regions automatically. We will see that our quality measure allows us to effectively compare these different strategies, and we find that the best coupled solutions are obtained by judiciously chosen transmission conditions.

2 Methods Based on Overlap and Optimization

In this section, we present a very general overlapping coupling strategy that was proposed in [5], where the authors considered as the viscous model the incompressible Navier-Stokes equations, while the inviscid model was the potential equation (the assumption of a small vorticity is made).

For the model problem (1), the coupling strategy works as follows: in each subdomain, we solve the corresponding equation with a Dirichlet condition at the artificial

interface,

$$u_{ad}(x_0+\delta)=\lambda_1 \text{ and } \text{ if } a>0,\, u_a(x_0)=\lambda_2,$$

and then determine (λ_1,λ_2) to be a solution of the optimization problem

$$J(\lambda_1,\lambda_2) := \|u_{ad}-u_a\|^2_{L^2(x_0,x_0+\delta)} \longrightarrow \min.$$

The authors in [5] solve this optimization problem using a gradient type method, so that the adjoint equation also needs to be computed.

This coupling strategy based on optimization has been studied mathematically in [10] and [2] for our model problem in 2D, see also [6] for a complete description of the algorithms for the model problem, and also for the coupling of Navier-Stokes equations with a Darcy model, or the coupling of the Stokes and potential equations. In [2] other cost functionals to be minimized are proposed.

In order to evaluate the quality of this coupling strategy, we compute numerically the error between the viscous and the coupled solution as a function of the viscosity for the case $L_1=L_2=1, x_0=-0.6, f(x)=e^{-1{,}000(x+1)^2}$ and $c=1$. We use a centered finite difference scheme to discretize the two differential operators, with mesh size 2×10^{-5}. We consider the case of a positive velocity, $a=1$, with $g_1=0, g_2=0, \mathscr{B}_1=Id$ and $\mathscr{B}_2=\partial_x-(a-\sqrt{a^2+4\nu c})/2\nu$ (the absorbing boundary operator) and the case of a negative velocity, $a=-1$, with $g_1=0, g_2=0, \mathscr{B}_1=Id$ and $\mathscr{B}_2=Id$. In all experiments presented in this paper, the error in the advection domain $\|u-u_a\|_{\Omega^+}$ is $\mathscr{O}(\nu)$ whatever is the coupling strategy, which is natural, since the advection equation is used instead of the advection-diffusion equation. The numerical error estimate for this overlapping technique in the viscous domain Ω^- is given in Table 1. We see that

	$a>0$	$a<0$
Minimization of J	$\mathscr{O}(\nu^{3/2})$	$\mathscr{O}(\nu)$

Table 1. Overlapping coupling with optimization: numerically computed error estimate for $\|u-u_{ad}\|_{\Omega^-}$

for $a<0$, this coupling strategy (like most of the ones presented in this paper) gives a result $\mathscr{O}(\nu)$, since information is coming from the inviscid approximation in Ω^+ to Ω^-, and in Ω^+ the error $\|u-u_a\|_{\Omega^+}$ is $\mathscr{O}(\nu)$.

The non overlapping case $\delta=0$ is also considered in [10], namely

$$G(\lambda_1,\lambda_2)=\sigma(a)(u_{ad}(x_0)-u_a(x_0))^2+(\phi_1-\phi_2)^2,$$

where $\phi_1=-\nu u'_{ad}(x_0)+au_{ad}(x_0)$ and $\phi_2=au_a(x_0)$ (see Sect. 3.1) and $\sigma(a)=1$ if $a>0$, 0 otherwise. Using the same numerical setting, we obtain for ν small the error estimates shown in Table 2.

	$a>0$	$a<0$
Minimization of G	$\mathcal{O}(\nu^{3/2})$	$\mathcal{O}(\nu)$

Table 2. Non overlapping case with optimization: numerically computed error estimates for $\|u-u_{ad}\|_{\Omega^-}$

3 Methods Based on Coupling Conditions

From now on we assume that there is no overlap, $\delta = 0$. The coupling techniques in this section are based on coupling conditions, and we will present three strategies: the first one is based on singular perturbation, the second one on boundary layer corrections, and the last one on the factorization of the operator.

3.1 Coupling Conditions from Singular Perturbation

In [9] the authors propose to find coupling conditions for (2) and (3) by introducing a regularization of the inviscid problem using a small artificial viscosity ε. They thus consider

$$\begin{aligned} -\nu w_\varepsilon'' + a w_\varepsilon' + c w_\varepsilon &= f \quad \text{on } (-L_1, x_0), \\ -\varepsilon v_\varepsilon'' + a v_\varepsilon' + c v_\varepsilon &= f \quad \text{on } (x_0, L_2). \end{aligned} \tag{4}$$

This coupling problem which involves two elliptic equations needs to be completed by two boundary conditions. The first one simply states continuity of the solution: $w_\varepsilon(x_0) = v_\varepsilon(x_0)$. For the second one, two choices are possible : we can impose the continuity of the normal flux, $\nu w_\varepsilon{}'(x_0) = \varepsilon v_\varepsilon{}'(x_0)$ (such boundary conditions are called variational conditions) or we impose the continuity of the normal derivative, $w_\varepsilon{}'(x_0) = v_\varepsilon{}'(x_0)$ (called non variational conditions). Letting ε tend to 0, it has been rigorously proved in [9] that w_ε (resp. v_ε) tends to u_{ad} (resp. u_a). At the boundary, with the variational conditions, the limiting solution satisfies

$$\begin{aligned} &(-\nu u_{ad}' + a u_{ad})(x_0) = a u_a(x_0), \quad u_{ad}(x_0) = u_a(x_0) \quad \text{for} \quad a > 0, \\ &(-\nu u_{ad}' + a u_{ad})(x_0) = a u_a(x_0), \qquad\qquad\qquad\quad \text{for} \quad a < 0, \end{aligned} \tag{5}$$

while the non variational conditions lead to

$$\begin{aligned} &u_{ad}(x_0) = u_a(x_0), \quad u_{ad}'(x_0) = u_a'(x_0), \quad \text{for} \quad a > 0, \\ &u_{ad}(x_0) = u_a(x_0), \qquad\qquad\qquad\quad \text{for} \quad a < 0. \end{aligned} \tag{6}$$

Rigorous error estimates comparing the coupled solutions obtained with these approaches were obtained in [7], and they are summarized in Table 3, where we observe that the non variational conditions lead to a better coupled solution for positive advection than the variational ones, while for negative advection, again there is no difference between the two approaches. Finally, it has been proved in [6] that the coupling problem with variational conditions is equivalent to the problem using optimization on $\sigma(a)(u_{ad}(0) - u_a(0))^2 + (\phi_1 - \phi_2)^2$; our observation is thus consistent.

	$a>0$	$a<0$
Variational Conditions	$\mathcal{O}(\nu^{3/2})$	$\mathcal{O}(\nu)$
Non Variational Conditions	$\mathcal{O}(\nu^{5/2})$	$\mathcal{O}(\nu)$

Table 3. Variational versus non-variational coupling conditions: theoretical error estimates for $\|u-u_{ad}\|_{\Omega^-}$

3.2 Coupling Through Boundary Layer Correction

A different approach, only adding a correction for the boundary layer (in the case $a<0$), was proposed in [4]. Here, the authors define the coupled solution of interest to be the solution of the regularized problem (4), and they consider the variational solution obtained from (5) as a first approximation of the regularized one. More precisely the coupled solution is represented as a perturbation of the variational solution in the form

$$\begin{aligned} w_\varepsilon(x) &= u_{ad}(x)+r_\varepsilon(x), \\ v_\varepsilon(x) &= u_a(x)+l_\varepsilon(x)+s_\varepsilon(x), \end{aligned}$$

where l_ε is a boundary layer function and r_ε and s_ε are the remainders of the asymptotic expansion. The boundary layer term can be computed analytically, but integrals that are involved are then approximated numerically. The numerical solution does not take into account the remainders r_ε and s_ε and thus, compared to the solution obtained with (5), the pure advection solution in Ω^+ is the only one to be corrected.

3.3 Coupling Conditions from Operator Factorization

A very accurate set of coupling conditions can be derived from an operator factorization, see [7], and requires the solution of a modified advection equation: if we introduce $\lambda^\pm=(a\pm\sqrt{a^2+4\nu c})/2\nu$, the advection diffusion equation can be factored, i.e.

$$\mathscr{L}_{ad}u=(\partial_x-\lambda^+)(\partial_x-\lambda^-)u=f,$$

which gives after integration on (x_0,L_2)

$$(\partial_x-\lambda^-)u(x_0)=(\partial_x-\lambda^-)u(L_2)e^{-\lambda^+L_2}+\int_{x_0}^{L_2}f(\sigma)e^{-\lambda^+\sigma}d\sigma.$$

Introducing the new advection equation $(\partial_x-\lambda^+)\tilde{u}_a=f$, we find that the viscous solution satisfies

$$(\partial_x-\lambda^-)u(x_0)=\tilde{u}_a(x_0)+((\partial_x-\lambda^-)u(L_2)-\tilde{u}_a(L_2))e^{-\lambda^+L_2}. \tag{7}$$

Solving the advection-diffusion equation in Ω^- with the boundary condition (7) (replacing u by u_{ad} on the left hand side) would thus yield the exact coupled solution, i.e. $u_{|\Omega^-}=u_{ad}$. However the term in L_2 can not be used directly, and one chooses instead $\tilde{u}_a(L_2)$ to be an expansion of $(\partial_x-\lambda^-)u(L_2)$ for ν small, so that the proposed coupling condition is

$$(\partial_x - \lambda^-)u_{ad}(x_0) = \tilde{u}_a(x_0). \tag{8}$$

This leads to the coupling procedure

1. Solve the new advection equation $(\partial_x - \lambda^+)\tilde{u}_a = f$ on (x_0, L_2) with $\tilde{u}_a(L_2) = z_0 + z_1\nu + \cdots + \mathcal{O}(\nu^m)$.
2. Solve the advection-diffusion equation on $(-L_1, x_0)$ with the transmission condition (8).
3. Solve the advection equation (3) on (x_0, L_2) with the condition $u_{ad}(x_0) = u_a(x_0)$ if $a > 0$.

For our model problem, rigorous error estimates obtained in [7] are shown in Table 4. We see that this coupling strategy leads to a coupled solution which is much closer to the fully viscous one than any of the other strategies. Even in the case of negative advection, one can now obtain approximations more accurate than $\mathcal{O}(\nu)$. Note however that $\lambda^\pm$ are simple constants only in the stationary one dimensional case. In the case of evolution, or for higher dimensions, the $\lambda^\pm$ need to be approximated (see for example [8]).

4 The χ-Formulation

A very different approach for coupling viscous and inviscid problems is proposed in [3]: the method called χ-formulation decides automatically where the viscous model and where the inviscid one needs to be used, and solves the equation

$$\begin{aligned} -\nu\chi(u'') + au' + cu &= f \quad \text{on } (-L_1, L_2), \\ u &= g_1 \text{ on } x = -L_1, \\ \mathcal{B}u &= 0 \quad \text{on } x = L_2, \end{aligned}$$

where the χ function is defined by

$$\chi(s) = \begin{cases} 0 & 0 \le s < \delta - \sigma, \\ (s - \delta + \sigma)\frac{\delta}{\sigma} & \delta - \sigma \le s \le \delta, \\ s & s > \delta, \end{cases}$$

so that the diffusion term is neglected as soon as it is small enough. This leads however to a non-linear equation, even if the underlying models are linear, which requires a Newton type algorithm.

In [3], the method is studied for the model problem at the continuous level, and well posedness is proved. Several years later, in [1] and [11], this strategy is used to solve the Navier-Stokes equations. Note that other cut-off functions can also be considered. We show in Table 5 numerically computed error estimates for the χ-formulation applied to our model problem.

	$a > 0$	$a < 0$
Factorization of the operator	$\mathcal{O}(e^{-a/\nu})$	$\mathcal{O}(\nu^m)$

Table 4. Coupling based on factorization: theoretical error estimates for $\|u - u_{ad}\|_{\Omega^-}$

	$a>0$	$a<0$
χ-formulation	$\mathcal{O}(\nu^{5/2})$	$\mathcal{O}(\nu)$

Table 5. χ-formulation: numerically computed error estimate for $\|u-u_{ad}\|_{\Omega^-}$

5 Conclusions

For a positive velocity a, among all the strategies presented in this paper, the best coupling condition is provided by the factorization of the operator in the non overlapping case: the error between the corresponding coupled solution and the fully viscous solution is exponentially small. Note that in the unstationary case or in higher dimensions the exponential convergence will be replaced by a polynomial one, because of approximations, an issue we currently investigate. Good algebraically small errors of $\mathcal{O}(\nu^{5/2})$ can also be obtained using the non variational conditions (6), or with the χ-formulation. The other strategies yield less accurate error estimates. When $a<0$, the factorization method is the only one to provide a better estimate than $\mathcal{O}(\nu)$.

Bibliography

[1] Y. Achdou and O. Pironneau. The χ-method for the Navier-Stokes equations. *IMA J. Numer. Anal.*, 13(4):537–558, 1993.

[2] V. Agoshkov, P. Gervasio, and A. Quarteroni. Optimal control in heterogeneous domain decomposition methods for advection-diffusion equations. *Mediterr. J. Math.*, 3(2):147–176, 2006.

[3] F. Brezzi, C. Canuto, and A. Russo. A self-adaptive formulation for the Euler/Navier-Stokes coupling. *Comput. Methods Appl. Mech. Engrg.*, 73(3): 317–330, 1989.

[4] C. A. Coclici, G. Moroşanu, and W. L. Wendland. The coupling of hyperbolic and elliptic boundary value problems with variable coefficients. *Math. Methods Appl. Sci.*, 23(5):401–440, 2000.

[5] Q. V. Dinh, R. Glowinski, J. Périaux, and G. Terrasson. On the coupling of viscous and inviscid models for incompressible fluid flows via domain decomposition. In *First International Symposium on Domain Decomposition Methods for Partial Differential Equations (Paris, 1987)*, pages 350–369. SIAM, Philadelphia, PA, 1988.

[6] M. Discacciati, Gervasio P., and A. Quarteroni. Heterogeneous mathematical models in fluid dynamics and associated solution algorithms. *Tech. Report MOX 04/2010*, 2010.

[7] M. J. Gander, L. Halpern, C. Japhet, and V. Martin. Viscous problems with inviscid approximations in subregions: a new approach based on operator factorization. In *CANUM 2008*, volume 27 of *ESAIM Proc.*, pages 272–288. EDP Sci., Les Ulis, 2009.

[8] M. J. Gander, L. Halpern, and V. Martin. How close to the fully viscous solution can one get with inviscid approximations in subregions ? In *Domain*

Decomposition Methods in Science and Engineering XIX, volume 78 of *Lect. Notes Comput. Sci. Eng.*, pages 237–244. Springer, 2011.

[9] F. Gastaldi and A. Quarteroni. On the coupling of hyperbolic and parabolic systems: analytical and numerical approach. *Appl. Numer. Math.*, 6(1–2):3–31, 1989/90. Spectral multi-domain methods (Paris, 1988).

[10] P. Gervasio, J.-L. Lions, and A. Quarteroni. Heterogeneous coupling by virtual control methods. *Numer. Math.*, 90(2):241–264, 2001.

[11] C.-H. Lai, A. M. Cuffe, and K. A. Pericleous. A defect equation approach for the coupling of subdomains in domain decomposition methods. *Comput. Math. Appl.*, 35(6):81–94, 1998.

Coupling Geometrically Exact Cosserat Rods and Linear Elastic Continua

Oliver Sander

Freie Universität Berlin, Fachbereich Mathematik und Informatik, Arnimallee 6, Berlin.
sander@mi.fu-berlin.de

Summary. We consider the mechanical coupling of a geometrically exact Cosserat rod to a linear elastic continuum. The coupling conditions are formulated in the nonlinear rod configuration space. We describe a Dirichlet–Neumann algorithm for the coupled system, and use it to simulate the static stresses in a human knee joint, where the Cosserat rods are models for the ligaments.

1 Cosserat Rods and Linear Elasticity

Cosserat rods are models for long slender objects. Let $\mathrm{SE}(3) = \mathbb{R}^3 \rtimes \mathrm{SO}(3)$ be the group of orientation-preserving rigid body motions of $\mathbb{R}^3$ (the special Euclidean group). A configuration of a Cosserat rod is a map $\varphi : [0,1] \to \mathrm{SE}(3)$. For each $s \in [0,1]$, the value $\varphi(s) = (\varphi_r(s), \varphi_q(s))$ is interpreted as the position $\varphi_r(s) \in \mathbb{R}^3$ and orientation $\varphi_q(s) \in \mathrm{SO}(3)$ of a rigid rod cross section. Strain measures $(\mathbf{v}_\varphi(s), \mathbf{u}_\varphi(s))$ at $\varphi(s)$ live in the tangent space $T_{\varphi(s)}\mathrm{SE}(3)$, and are defined by

$$\mathbf{v}_\varphi(s) = \varphi_r'(s) \qquad \text{and} \qquad \varphi_q'(s) = \mathbf{u}_\varphi^\times(s)\varphi_q(s),$$

where $\mathbf{u}_\varphi^\times$ is the skew-symmetric matrix corresponding to $\mathbf{u}_\varphi$. On each cross section s of the rod act a resultant force and torque. These are given by a tuple $(\mathbf{n}(s), \mathbf{m}(s))$, which is an element of the cotangent space $T^*_{\varphi(s)}\mathrm{SE}(3)$. In the absence of external forces and torques we have the equations of equilibrium [6]

$$\begin{aligned} \mathbf{m}' + \varphi_r' \times \mathbf{n} &= 0 \qquad \text{on } [0,1], \\ \mathbf{n}' &= 0 \qquad \text{on } [0,1]. \end{aligned}$$

We assume there to be an energy functional W such that $\mathbf{n} = \partial W/\partial \mathbf{v}$ and $\mathbf{m} = \partial W/\partial \mathbf{u}$. Existence of solutions for this model has been shown in [12], but note that solutions may be nonunique.

We will couple the rod model to a linear elastic continuum. Let Ω be a domain in $\mathbb{R}^3$. Its boundary $\partial\Omega$ is supposed to be Lipschitz and to consist of disjoint parts

R. Bank et al. (eds.), *Domain Decomposition Methods in Science and Engineering XX*, Lecture Notes in Computational Science and Engineering 91,
DOI 10.1007/978-3-642-35275-1_52,

Γ_N and Γ_D such that $\partial\Omega = \overline{\Gamma}_N \cup \overline{\Gamma}_D$ and Γ_D has positive two-dimensional measure. We use $\boldsymbol{\nu}_\Omega$ to denote the outward unit normal of Ω. For any displacement function $\mathbf{u} \in \mathbf{H}^1(\Omega) = (H^1(\Omega))^3$ we set $\boldsymbol{\varepsilon} = \frac{1}{2}(\nabla\mathbf{u} + \nabla\mathbf{u}^T)$ the linear strain tensor and the stress $\boldsymbol{\sigma} = \boldsymbol{\sigma}(\boldsymbol{\varepsilon})$, with a St. Venant–Kirchhoff-type material law

$$\boldsymbol{\sigma}(\boldsymbol{\varepsilon}) = \frac{E\nu}{(1+\nu)(1-2\nu)}(\operatorname{tr}\boldsymbol{\varepsilon})\mathrm{Id} + \frac{E}{1+\nu}\boldsymbol{\varepsilon}.$$

The parameters E and ν are the Young's modulus and Poisson ratio, respectively. The boundary value problem of elasticity is then

$$\begin{aligned} -\operatorname{div}\boldsymbol{\sigma}(\mathbf{u}) &= \mathbf{f} && \text{in } \Omega, \\ \mathbf{u} &= 0 && \text{on } \Gamma_D, \\ \boldsymbol{\sigma}(\mathbf{u})\boldsymbol{\nu}_\Omega &= \mathbf{t} && \text{on } \Gamma_N, \end{aligned}$$

with volume forces $\mathbf{f} : \Omega \to \mathbb{R}^3$ and surface force $\mathbf{t} : \Gamma_N \to \mathbb{R}^3$.

2 Coupling Conditions

We will now derive conditions for the coupling of a Cosserat rod and a linear elastic three-dimensional object. The two main difficulties are the difference in dimensions between the rod and the continuum, and the nonlinear nature of the rod configuration space.

Previous work has mainly focused on coupling linear models of different dimensions. Lagnese et al. [7] have studied the coupling of beams to plates extensively. Modeling of 3d–2d junctions between linear elastic objects using a method of asymptotic expansion has been carried out by Ciarlet et al. [4]. Monaghan et al. [8] describe a 3d–1d coupling between linear elastic elements in the discrete setting. A general framework which encompasses these cases is given in [3]. We are not aware of previous work on the coupling of Cosserat rods.

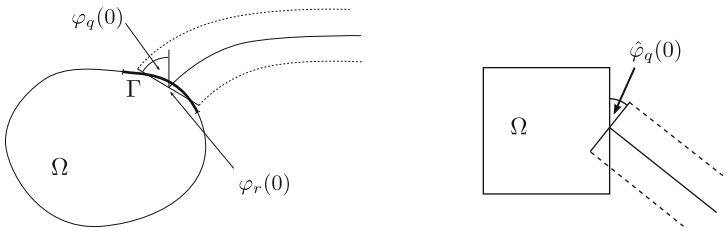


Fig. 1. *Left*: Coupling between a two-dimensional domain and a rod. *Right*: In the stress-free configuration the rod may meet the body at an arbitrary spatial angle $\hat{\varphi}_q(0)$

Consider again a linear elastic continuum defined on a reference configuration Ω. This time, the boundary $\partial\Omega$ is supposed to consist of three disjoint parts Γ_D, Γ_N,

and Γ such that $\partial\Omega = \overline{\Gamma}_D \cup \overline{\Gamma}_N \cup \overline{\Gamma}$. We assume that Γ_D and Γ have positive two-dimensional measure. The three-dimensional object represented by Ω will couple with the rod across Γ, which we call the coupling boundary. The boundary of the parameter domain $[0,1]$ of a Cosserat rod consists only of the two points 0 and 1, and the respective domain normals are $\boldsymbol{\nu}_{r,0} = -1$ and $\boldsymbol{\nu}_{r,1} = 1$. To be specific, we pick 0 as the coupling boundary. We assume a stress-free rod configuration $\hat{\varphi} : [0,1] \to \mathrm{SE}(3)$ such that $\hat{\varphi}_r(0) = |\Gamma|^{-1} \int_\Gamma x\, ds$, i.e., the coupling interface of the rod in its stress-free state is placed at the center of gravity of the coupling interface of Ω. The orientation $\hat{\varphi}_q(0)$ of the stress-free state does not need to be in any relation with the shape of the coupling boundary Γ (Fig. 1).

We define our coupling using a set of conditions for the primal variables. These variables are the configuration φ of the rod and the displacement field $\mathbf{u}$ of the continuum. It is well known that when coupling two continuum models of the same type, the solution has to be continuous [9]. Since the position $\varphi_r(0) \in \mathbb{R}^3$ of the coupling cross-section can be seen as an averaged position it is natural to couple it to the averaged position of Γ

$$\varphi_r(0) \stackrel{!}{=} \frac{1}{|\Gamma|} \int_\Gamma (\mathbf{u}(x) + x)\, ds. \tag{1}$$

To obtain a complete set of primal conditions we also need to relate the orientations at the interface. This requires some technical preparations. Using the deformation gradient $F(\mathbf{u}) = \boldsymbol{\nabla}(\mathbf{u} + \mathrm{Id})$ we first define the average deformation of the interface boundary Γ as $\mathscr{F}(\mathbf{u}) = |\Gamma|^{-1} \int_\Gamma \boldsymbol{\nabla}(\mathbf{u}(x) + x)\, ds$. If $\mathbf{u}$ stays within the limits of linear elasticity the matrix $\mathscr{F}(\mathbf{u})$ has a positive determinant. Using the polar decomposition it can then be split into a rotation $\mathrm{polar}(\mathscr{F}(\mathbf{u}))$ and a stretching. We define the average orientation of Γ induced by a deformation $\mathbf{u}$ as the rotational part of $\mathscr{F}(\mathbf{u})$. This corresponds to the definition of the continuum rotation used in the theory of Cosserat continua. In particular, if $\mathbf{u} \equiv 0$ then $\mathrm{polar}(\mathscr{F}(\mathbf{u})) = \mathrm{Id}$.

The average orientation $\mathrm{polar}(\mathscr{F}(\mathbf{u}))$ can now be set in relation to $\varphi_q(0)$, the orientation of the rod cross-section at $s = 0$. We require the coupling condition to be fulfilled by the stress-free configuration $\mathbf{u} = 0$, $\varphi = \hat{\varphi}$. This leads to the condition

$$\varphi_q(0) \stackrel{!}{=} \mathrm{polar}(\mathscr{F}(\mathbf{u}))\hat{\varphi}_q(0), \tag{2}$$

which is an equation in the nonlinear three-dimensional space SO(3).

For ease of writing we will introduce the averaging operator $\mathrm{Av} : \mathbf{H}^1(\Omega) \to \mathrm{SE}(3)$ by setting

$$\mathrm{Av}(\mathbf{u}) = \left(\frac{1}{|\Gamma|} \int_\Gamma (\mathbf{u}(x) + x)\, ds,\, \mathrm{polar}(\mathscr{F}(\mathbf{u}))\hat{\varphi}_q(0) \right), \tag{3}$$

where we have used $(\cdot,\cdot)$ to denote elements of the product space $\mathrm{SE}(3) = \mathbb{R}^3 \rtimes \mathrm{SO}(3)$. It is a nonlinear generalization of the restriction operator used in [3]. Then (1) and (2) can be written concisely as

$$\varphi(0) \stackrel{!}{=} \mathrm{Av}(\mathbf{u}). \tag{4}$$

Note that we do not assume that Γ has the same shape or area as the rod cross-section at $s = 0$. Also, since the coupling conditions relate only finite-dimensional quantities

they remain the same when the subdomain problems are replaced by finite element approximations.

The coupling problem is made complete by conditions for the dual variables. For the continuum these variables are the normal stresses at the boundary Γ. For the rod the dual variables are the total force $\mathbf{n}(0)\boldsymbol{\nu}_{r,0}$ and the total moment $\mathbf{m}(0)\boldsymbol{\nu}_{r,0}$ about $\varphi_r(0)$ transmitted in normal direction across the cross-section at $s = 0$. We expect these to match the total force and torque exerted by the continuum across the coupling boundary Γ in the direction of $-\boldsymbol{\nu}_\Omega$

$$\int_\Gamma \boldsymbol{\sigma}(\mathbf{u})\boldsymbol{\nu}_\Omega \, ds = -\mathbf{n}(0)\boldsymbol{\nu}_{r,0} \tag{5}$$

$$\int_\Gamma (x - \varphi_r(0)) \times (\boldsymbol{\sigma}(\mathbf{u})\boldsymbol{\nu}_\Omega) \, ds = -\mathbf{m}(0)\boldsymbol{\nu}_{r,0}. \tag{6}$$

Together, these equations relate quantities in the six-dimensional space $T^*_{\varphi(0)}\mathrm{SE}(3)$.

Remark 1. A variational formulation suggests that (5) and (6) are not the dual conditions of (4) (cf. to [3] for the linear case). Together with (10), however, they are sufficient to construct a working solution algorithm.

3 A Dirichlet–Neumann Algorithm

In this section we present a Dirichlet–Neumann algorithm for the coupled problem. It can be interpreted as a fixed-point iteration for an equation on the trace space of the rod configuration space at $s = 0$, i.e. on SE(3). Each iteration consists of three steps: a Dirichlet problem for the rod, a Neumann problem for the body, and a damped update along geodesics on SE(3). Let $\lambda^0 \in \mathrm{SE}(3)$ be the initial interface value and $k \geq 0$ the iteration number. In more detail, the steps are as follows.

1. *Dirichlet problem for the Cosserat rod*
   Let $\lambda^k, \varphi_D \in \mathrm{SE}(3)$ be the current interface value and a Dirichlet boundary value, respectively. Find a solution φ^{k+1} of the Dirichlet rod problem

$$\begin{aligned}
(\mathbf{m}^{k+1})' + (\varphi_r^{k+1})' \times \mathbf{n}^{k+1} &= 0 && \text{on } [0,1] \\
(\mathbf{n}^{k+1})' &= 0 && \text{on } [0,1] \\
\varphi^{k+1}(0) &= \lambda^k \\
\varphi^{k+1}(1) &= \varphi_D.
\end{aligned}$$

2. *Neumann problem for the continuum*
   The new rod iterate φ^{k+1} exerts a resultant force $\mathbf{n}^{k+1}(0)\boldsymbol{\nu}_{r,0}$ and moment $\mathbf{m}^{k+1}(0)\boldsymbol{\nu}_{r,0}$ across its cross-section at $s = 0$. Construct a Neumann data field $\boldsymbol{\tau}^{k+1} : \Gamma \to \mathbb{R}^3$ such that

$$\int_\Gamma \boldsymbol{\tau}^{k+1}(x) \, ds = -\mathbf{n}^{k+1}(0)\boldsymbol{\nu}_{r,0} \tag{7}$$

and

$$\int_\Gamma (x - \varphi_r^{k+1}(0)) \times \boldsymbol{\tau}^{k+1}(x)\, ds = -\mathbf{m}^{k+1}(0)\boldsymbol{\nu}_{r,0}. \tag{8}$$

Then solve the three-dimensional linear elasticity problem with Neumann data $\boldsymbol{\tau}^{k+1}$ on Γ

$$\begin{aligned} -\operatorname{div}\boldsymbol{\sigma}(\mathbf{u}^{k+1}) &= \mathbf{f} && \text{in } \Omega \\ \boldsymbol{\sigma}(\mathbf{u}^{k+1})\boldsymbol{\nu}_\Omega &= \boldsymbol{\tau}^{k+1} && \text{on } \Gamma \\ \mathbf{u}^{k+1} &= 0 && \text{on } \Gamma_D \\ \boldsymbol{\sigma}(\mathbf{u}^{k+1})\boldsymbol{\nu}_\Omega &= \mathbf{t} && \text{on } \Gamma_N. \end{aligned} \tag{9}$$

3. *Damped geodesic update*
From the solution $\mathbf{u}^{k+1}$ compute the average interface displacement and orientation $\operatorname{Av}(\mathbf{u}^{k+1})$ as defined in (3). With a damping parameter $\theta > 0$, the new interface value λ^{k+1} is then computed as a geodesic combination in SE(3) of the old value λ^k and $\operatorname{Av}(\mathbf{u}^{k+1})$,

$$\lambda^{k+1} = \exp_{\lambda^k} \theta \left[\exp_{\lambda^k}^{-1} \operatorname{Av}(\mathbf{u}^{k+1}) \right].$$

It remains to say how to construct suitable fields of Neumann data $\boldsymbol{\tau}^{k+1}$ that satisfy the conditions (7) and (8). Let us drop the index k for simplicity. In principle, any function $\boldsymbol{\tau} : \Gamma \to \mathbb{R}^3$ of sufficient regularity fulfilling (7) and (8) can be used as Neumann data in (9). It has been shown in [10] that such functions exist.

The theory of Cosserat rods assumes that forces and moments are transmitted evenly across cross-sections. We therefore construct $\boldsymbol{\tau}$ to be 'as constant as possible'. More formally, we introduce the functional

$$T : \mathbf{L}^2(\Gamma) \times \mathbb{R}^3 \to \mathbb{R}, \qquad T(\mathbf{h}, \mathbf{c}) = \int_\Gamma \|\mathbf{h}(x) - \mathbf{c}\|^2 ds,$$

and construct $\boldsymbol{\tau}$ as the solution of the minimization problem

$$(\boldsymbol{\tau}, \mathbf{c}_{\boldsymbol{\tau}}) = \operatorname*{arg\,min}_{\mathbf{h} \in \mathbf{L}^2(\Gamma), \mathbf{c} \in \mathbb{R}^3} T(\mathbf{h}, \mathbf{c}) \tag{10}$$

under the constraints that

$$\int_\Gamma \boldsymbol{\tau}\, ds = -\mathbf{n}(0)\boldsymbol{\nu}_{r,0} \qquad \text{and} \qquad \int_\Gamma (x - \varphi_r(0)) \times \boldsymbol{\tau}\, ds = -\mathbf{m}(0)\boldsymbol{\nu}_{r,0}. \tag{11}$$

Problem (10) and (11) is a convex minimization problem with linear equality constraints. In [10, Lemma 5.3.4] it was shown that there exists a unique solution. In a finite element setting the problem size is given by the number of grid vertices on Γ times 3. A minimization problem of this type can be solved, e.g., with an interior-point method.

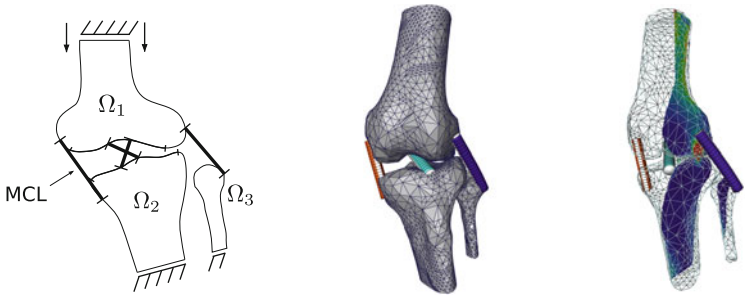


Fig. 2. *Left*: Problem setting. Tibia and fibula are rotated 15° in valgus direction to put additional stress on the MCL. *Center*: Deformed grids after two adaptive refinement steps. *Right*: Two sagittal cuts through the von Mises stress field

4 Numerical Results

We close with a simulation result for a knee model which combines femur, tibia, and fibula bones modeled as three-dimensional linear elastic objects, and the cruciate and collateral ligaments, modeled as Cosserat rods. The model additionally includes the contact between femur and tibia. To obtain a test case where the contact stresses do not entirely predominate the stresses created in the bone by pulling ligaments, we applied a valgus rotation of 15° to tibia and fibula. This leads to a high strain in the medial collateral ligament (MCL) and can be interpreted as an imminent MCL rupture (Fig. 2).

The geometry was obtained from the Visible Human data set. We modeled bone with an isotropic, homogeneous, linear elastic material with $E = 17\,\text{GPa}$ and $\nu = 0.3$. The distal horizontal sections of tibia and fibula were clamped, and a prescribed downward displacement of 2 mm was applied to the upper section of the femur. We used first-order finite elements for the discretization of the linear elasticity problem. DUNE [2] was used for the implementation.

The four ligaments were each modeled by a single Cosserat rod with a circular cross-section of radius 5 mm. The rod equations were discretized using geodesic finite elements [11]. We chose a linear material law (see, e.g., [6]) with parameters $E = 330\,\text{MPa}$ and $\nu = 0.3$. On the bones, the coupling boundaries Γ for the different ligaments were marked by hand using a graphical editor. We modeled all ligaments to be straight in their stress-free configurations and to have 8 % in situ strain.

We solved the combined problem using the Dirichlet–Neumann algorithm described in Sect. 3. At each iteration, a pure Dirichlet problem had to be solved for each of the rods and a contact problem with mixed Dirichlet–Neumann boundary conditions had to be solved for the bones. The contact problem was solved using the Truncated Nonsmooth Newton Multigrid (TNNMG) algorithm [5]. The TNNMG method solves linear contact problems with the efficiency of linear multigrid. For the ligaments we used a Riemannian trust-region solver [1, 11], and we used IPOpt [13] to solve the minimization problems (10) and (11). Figure 2 shows the deformed con-

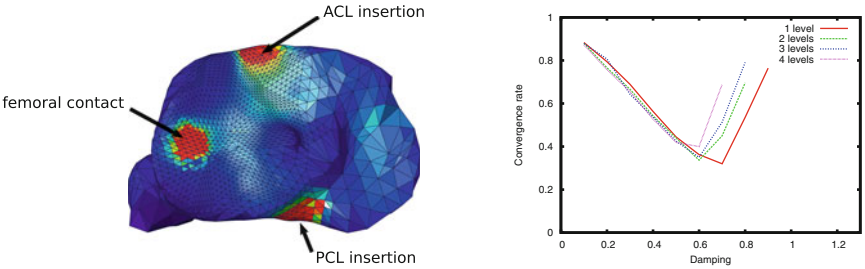


Fig. 3. *Left*: Stress plot on the tibial plateau. *Right*: Convergence rates of the Dirichlet–Neumann method as a function of the damping parameter for up to four grid levels

figuration on a grid obtained by two steps of adaptive refinement and cuts through the von Mises stress field. In Fig. 3, left, a caudal view onto the tibial plateau can be seen, which is colored according to the von Mises stress. The peaks due to contact and the pull of the cruciate ligaments can be clearly observed.

We measured the Dirichlet–Neumann convergence rates with bone grids obtained by up to three steps of adaptive refinement using the hierarchical error estimator presented in [10]. Rod grids in turn were refined uniformly. On each new set of grids we started the computation from the reference configuration. That way identical initial iterates for all grid refinement levels were obtained. Details on the measuring setup can be found in [10]. Figure 3, right, shows the Dirichlet–Neumann convergence rates plotted as a function of the damping parameter θ for up to four levels of refinement. For each further level of refinement, the optimal convergence rate is slightly worse than for the previous, and obtained for a slightly lower damping parameter. This behavior seems typical for Dirichlet–Neumann methods. Nevertheless the optimal convergence rates stay around 0.4. This makes the algorithm well usable in practice.

Bibliography

[1] P.-A. Absil, C. G. Baker, and K. A. Gallivan. Trust-region methods on Riemannian manifolds. *Found. Comput. Math.*, 7(3):303–330, 2007.

[2] P. Bastian, M. Blatt, A. Dedner, C. Engwer, R. Klöfkorn, R. Kornhuber, M. Ohlberger, and O. Sander. A generic interface for parallel and adaptive scientific computing. Part II: Implementation and tests in DUNE. *Computing*, 82(2–3):121–138, 2008.

[3] P. J. Blanco, M. Discacciati, and A. Quarteroni. Modeling dimensionally-heterogenenous problems: analysis, approximation and applications. Technical Report 14, MATHICSE, 2010.

[4] P. G. Ciarlet, H. LeDret, and R. Nzengwa. Junctions between three-dimensional and two-dimensional linearly elastic structures. *J. Math. Pures Appl.*, 68:261–295, 1989.

[5] C. Gräser, U. Sack, and O. Sander. Truncated nonsmooth Newton multigrid methods for convex minimization problems. In *Proc. of DD18*, LNCSE, pages 129–136. Springer, 2009.

[6] S. Kehrbaum. *Hamiltonian Formulations of the Equilibrium Conditions Governing Elastic Rods: Qualitative Analysis and Effective Properties*. PhD thesis, University of Maryland, 1997.

[7] J. Lagnese, G. Leugering, and E. Schmidt. *Modeling, Analysis and Control of Dynamic Elastic Multi-Link Structures*. Birkhäuser, 1994.

[8] D. J. Monaghan, I. W. Doherty, D. M. Court, and C. G. Armstrong. Coupling 1D beams to 3D bodies. In *Proc. 7th Int. Meshing Roundtable*. Sandia National Laboratories, 1998.

[9] A. Quarteroni and A. Valli. *Domain Decomposition Methods for Partial Differential Equations*. Oxford Science Publications, 1999.

[10] O. Sander. *Multidimensional Coupling in a Human Knee Model*. PhD thesis, Freie Universität Berlin, 2008.

[11] O. Sander. Geodesic finite elements for Cosserat rods. *Int. J. Num. Meth. Eng.*, 82(13):1645–1670, 2010.

[12] T. I. Seidman and P. Wolfe. Equilibrium states of an elastic conducting rod in a magnetic field. *Arch. Rational Mech. Anal.*, 102:307–329, 1988.

[13] A. Wächter and L. T. Biegler. On the implementation of a primal–dual interior point filter line search algorithm for large-scale nonlinear programming. *Math. Progr.*, 106(1):25–57, 2006.

Parareal Schwarz Waveform Relaxation Methods

Martin J. Gander[1], Yao-Lin Jiang[2], Rong-Jian Li[3]

[1] Mathematics Section, University of Geneva, CH-1211, Geneva, Switzerland martin.gander@unige.ch
[2] Department of Mathematics Sciences, Xi'an Jiaotong University, Xi'an, Shaanxi 710049, China yljiang@mail.xjtu.edu.cn
[3] Department of Mathematics Sciences, Xi'an Jiaotong University, Xi'an, Shaanxi 710049, China rongjian.li@stu.xjtu.edu.cn
This work was in part supported by the International Science and Technology Cooperation Program of China under grant 2010DFA14700.

1 Introduction

Solving an evolution problem in parallel is naturally undertaken by trying to parallelize the algorithm in space, and then still follow a time stepping method from the initial time $t = 0$ to the final time $t = T$. This is especially easy to do when an explicit time stepping method is used, because in that case the time step for each component is only based on past, known data, and the time stepping can be performed in an embarrassingly parallel way. If one uses implicit time stepping however, one obtains a large system of coupled equations, and thus the linear or non-linear solver needs to be parallelized, e.g. using a domain decomposition method.

Over the last decades, people have however also tried to parallelize algorithms in the time direction. One example is Womble's algorithm [22], where the systems arising from an implicit time discretization are solved using an iterative method, and the iteration of the next time level is started, before the iteration on the current time level has converged. It is then possible to iterate several time levels simultaneously, but the possible gain using a parallel computer is only small, see for example [3].

A different approach to obtain small scale parallelism in time is to use predictor-corrector methods, where the prediction step and the correction step can be performed by two (or several) processors in parallel, if organized properly. An entire class of such methods has been proposed in [19], and good small scale parallelism can be achieved.

A third, very different approach are the waveform relaxation algorithms, invented in [15], which are based on a decomposition of the system to be solved into subsystems. An iteration is then used, which solves time dependent problems in each subsystem and communicates information at interfaces to neighboring subsystems to converge to the overall solution in space-time [12, 13]. Substantial progress has been made on such methods for evolution PDEs, see for example [5, 6, 14], and references

R. Bank et al. (eds.), *Domain Decomposition Methods in Science and Engineering XX*, Lecture Notes in Computational Science and Engineering 91,
DOI 10.1007/978-3-642-35275-1_53,

therein. If a multi-grid decomposition is used, instead of a domain decomposition, one obtains the so called parabolic multi-grid methods [11], which are also called multi-grid waveform relaxation methods. For further results, see [17, 21].

Finally, the last class of methods, which focuses entirely on the parallelization in the time direction, are based on shooting methods in time. A first historical step in this direction is [20], and for an early analysis see [2]. The newest algorithm in this class is the parareal algorithm, invented in [16]. For a complete historical overview of such methods, further references, and a precise convergence estimate of the parareal algorithm see [4, 9].

We propose here a space time parallel algorithm for solving evolution partial differential equations, and use as a model problem

$$\begin{aligned} \partial_t u &= \partial_{xx} u \quad \text{in } \Omega = (0,1)\times(0,T), \\ \mathscr{B}^- u(0,t) &= g_0(t) \;\; t \in (0,T), \\ \mathscr{B}^+ u(1,t) &= g_1(t) \;\; t \in (0,T), \\ u(x,0) &= u_0(x) \;\; x \in \Omega. \end{aligned} \tag{1}$$

Here $\mathscr{B}^{\pm}$ represent some boundary operators, like the identity for a Dirichlet condition, or a normal derivative for a Neumann condition. The algorithm is based on a decomposition of the space-time domain into space-time subdomains, as indicated in Fig. 1. In order to solve an evolution problem by only solving problems in small space-time domains, one has to iteratively calculate more and more accurate initial and boundary conditions for each space-time subdomain. The parareal Schwarz waveform relaxation algorithm does this by using a parareal approximation for the initial conditions, and a Schwarz waveform relaxation algorithm for the boundary conditions. For a different variant of combining a spatial and a time decomposition, see [18].

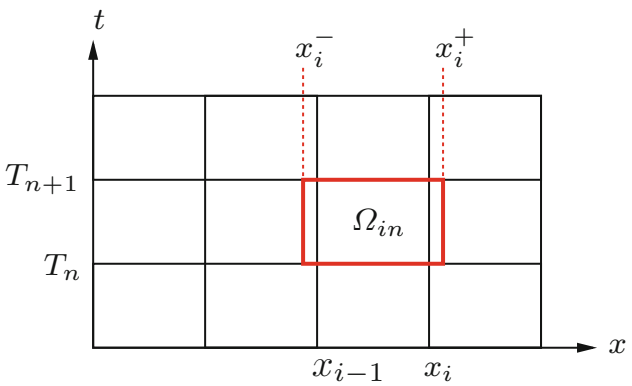


Fig. 1. Space time decomposition for the parareal Schwarz waveform relaxation algorithm

2 Parareal Schwarz Waveform Relaxation Algorithms

The parareal algorithm for the model problem (1) is based on a decomposition of the time interval $(0,T)$ into subintervals, given by $0 = T_0 < T_1 < T_2 < \ldots < T_N = T$, and the algorithm is defined using two propagation operators: a coarse operator $G(t_2,t_1,u_1,g_0,g_1)$ which provides a rough approximation of the solution $u(x,t_2)$ of (1) with a given initial condition $u(x,t_1) = u_1(x)$ and boundary conditions g_0 and g_1, and a fine operator $F(t_2,t_1,u_1,g_0,g_1)$, which gives a more accurate approximation of the same solution with initial condition $u(x,t_1) = u_1(x)$ and boundary conditions g_0 and g_1. Starting with a first approximation U_n^0 at the time points $T_0,T_1,T_2,\ldots,T_{N-1}$, the parareal algorithm performs for $k = 0,1,2,\ldots$ the correction iteration

$$U_{n+1}^{k+1} = F(T_{n+1},T_n,U_n^k,g_0,g_1) + G(T_{n+1},T_n,U_n^{k+1},g_0,g_1) - G(T_{n+1},T_n,U_n^k,g_0,g_1), \tag{2}$$

which is nothing else than a multiple shooting method with an approximate Jacobian in the Newton step, see for example [9], which also contains a precise convergence estimate for the case of the heat equation, or [4] for a similar precise convergence estimate for the case of nonlinear problems.

In contrast to the parareal algorithm, a Schwarz waveform relaxation method for the model problem (1) is based on a spatial decomposition only, in the most general case into overlapping subdomains $\Omega = \cup_{i=1}^{I}(x_i^-,x_i^+)$, as shown in Fig. 1. Here the boundaries $x_i^\pm$ of the overlapping subdomains are constructed from a non-overlapping decomposition given by the decomposition $0 =: x_0 < x_1 < \ldots < x_I := 1$, by adding and subtracting half the overlap, $x_i^- := x_{i-1} - \frac{L}{2}$, $x_i^+ := x_i + \frac{L}{2}$, except for the first and last point, $x_1^- = x_0$ and $x_I^+ = x_I$. Given an initial guess at the interfaces, say $\mathscr{B}_i^\pm u_i^0$, the Schwarz waveform relaxation algorithm solves iteratively for $k = 1,2,\ldots$ the subdomain problems

$$\begin{aligned}
\partial_t u_i^k &= \partial_{xx} u_i^k && \text{in } \Omega_i \times (0,T),\\
u_i^k(x,0) &= u_0 && \text{in } \Omega_i,\\
\mathscr{B}_i^- u_i^k(x_i^-,t) &= \mathscr{B}_i^- u_{i-1}^{k-1}(x_i^-,t) && t \in (0,T),\\
\mathscr{B}_i^+ u_i^k(x_i^+,t) &= \mathscr{B}_i^+ u_{i+1}^{k-1}(x_i^+,t) && t \in (0,T).
\end{aligned} \tag{3}$$

Here again, the operators $\mathscr{B}_i^\pm$ are transmission operators: in the case of the identity, we have the classical Schwarz waveform relaxation algorithm; for Robin or higher order transmission conditions, one would obtain an optimized Schwarz waveform relaxation algorithm, if the parameters in the transmission conditions are chosen to optimize the convergence of the algorithm, see [1, 5].

Parareal Schwarz waveform relaxation algorithms combine the two techniques for a general space-time decomposition given in Fig. 1. We propose among the many possibilities the following one: given initial conditions $u_{0,i,n}^k(x)$ and boundary conditions $\mathscr{B}_i^- u_{i-1,n}^k(t)$ and $\mathscr{B}_i^+ u_{i+1,n}^k(t)$ for $i = 1,2,\ldots,I$ and $n = 1,2,\ldots,N$ we compute

1. All accurate approximations $u_{i,n}^{k+1}(x,t) := F_{i,n}(u_{0,i,n}^k, \mathscr{B}_i^- u_{i-1,n}^k, \mathscr{B}_i^+ u_{i+1,n}^k)$ in parallel using the more accurate evolution operator.

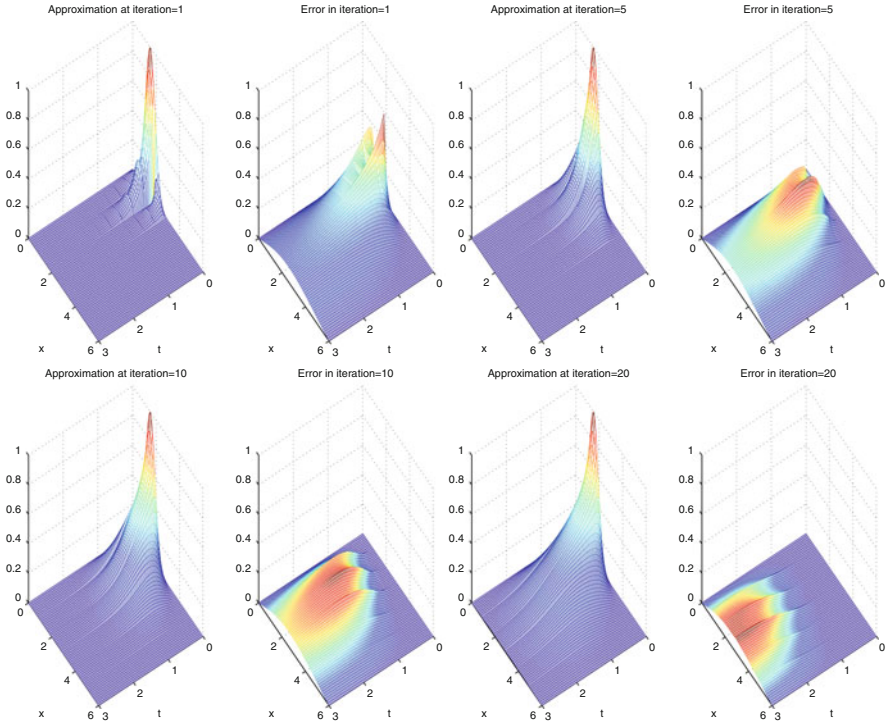


Fig. 2. Illustration how the parareal Schwarz waveform relaxation algorithm removes the error over several iterations: each plot pair shows on the *left* the approximation and on the *right* the error (i.e. the difference between the monodomain solution and the current iterate) for $k = 1, 5, 10, 20$

2. For $n = 0, 1, \ldots$, new initial conditions using a parareal integration step both in space and time,

$$u_{0,i,n+1}^{k+1} = u_{i,n}^{k+1}(\cdot, T_{n+1}) + G_{i,n}(u_{0,i,n}^{k+1}, \mathscr{B}_i^- u_{i-1,n}^{k+1}, \mathscr{B}_i^+ u_{i+1,n}^{k+1}) - G_{i,n}(u_{0,i,n}^{k}, \mathscr{B}_i^- u_{i-1,n}^{k}, \mathscr{B}_i^+ u_{i+1,n}^{k}).$$

An example on how this algorithm converges is given in Fig. 2.

We present now a first convergence result for the parareal Schwarz waveform relaxation algorithm:

Theorem 1 (Superlinear Convergence). *Let $F_{i,n}$ be the exact solution, $G_{i,n}$ be a backward Euler approximation in time, and the exact solution in space, and assume a decomposition of the spatial domain into two overlapping subdomains. If the algorithms uses Dirichlet transmission conditions, i.e. $\mathscr{B}_i^{\pm} = I$, the identity, then it converges superlinearly to the solution of the underlying problem.*

The proof of this theorem is too long and technical for this short paper, and will appear in [7]. We present however a detailed numerical study of how the algorithm depends on the various parameters in the following section.

3 Numerical Results

In all our experiments, except otherwise mentioned, we use the domain $\Omega = (0,6)$ and the time interval $(0,T)$ with $T = 3$, and discretize the heat equation with a centered finite difference discretization in space with $\Delta x = \frac{1}{10}$, and a backward Euler discretization in time, with $\Delta t = \frac{3}{100}$, and we use a decomposition into 6 equal spatial subdomains with overlap $2\Delta x$.

We start with the dependence on the number of time subintervals. In Fig. 3 on the left, we show the convergence of the algorithm when 1 (classical Schwarz waveform relaxation), 2, 4 and 10 time subintervals are used. This shows that the algorithm is quite insensitive to the number of time subintervals used. We also observe the typical superlinear convergence behavior of all waveform relaxation algorithms, see for example [8].

We next investigate how the convergence depends on the total time interval length T. For this experiment, leaving all other parameters the same, we choose $T \in \{0.1, 0.2, 0.4, 0.8, 1.6, 3.2\}, \Delta t = \frac{T}{100}$, and ten time subintervals for each simulation. The results are shown in Fig. 3 on the right. We clearly see that convergence is much faster on short time intervals, compared to long time intervals.

In order to test the dependence on the number of spatial subdomains, we use again all parameters as before, but now decompose the domain into 2, 3, 6 and 12 spatial subdomains, and again 10 time subintervals. We see in Fig. 4 on the left that using more spatial subdomains makes the algorithm converge more slowly. This can however be remedied by using smaller global time intervals, as for the Schwarz waveform relaxation algorithm, see [10].

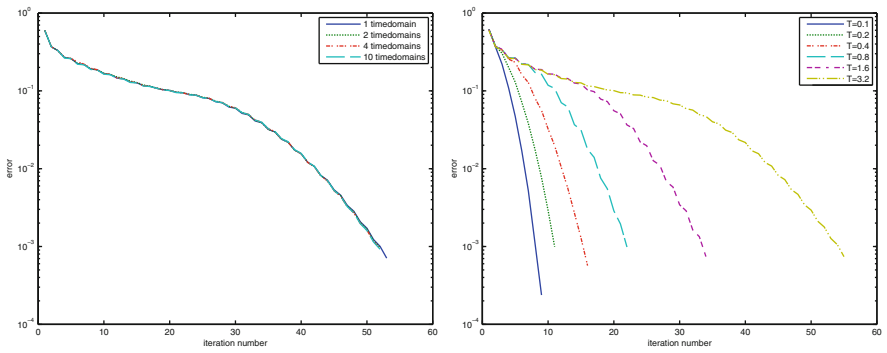


Fig. 3. Dependence of the parareal Schwarz waveform relaxation algorithm on the number of time subintervals on the *left*, and the total time window length on the *right*

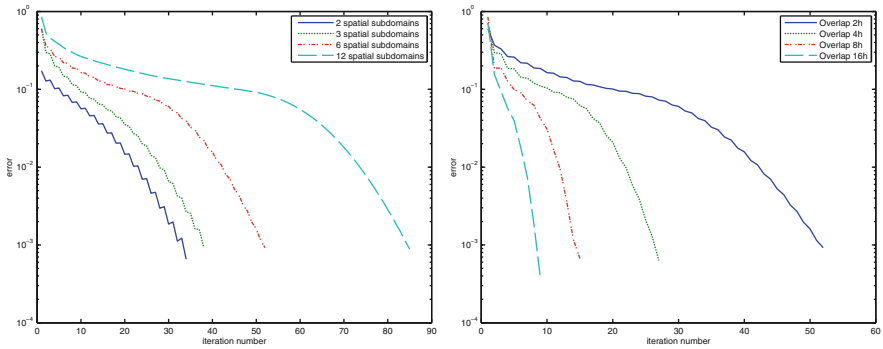


Fig. 4. Dependence of the parareal Schwarz waveform relaxation algorithm on the number of spatial subdomains on the *left*, and the overlap on the *right*

We finally test the dependence on the overlap, using $2\Delta x$, $4\Delta x$, $8\Delta x$ and $16\Delta x$ for the overlap. We see on the right in Fig. 4 that increasing the overlap substantially improves the convergence speed of the algorithm. This increases however also the cost of the method, since bigger subdomain problems need to be solved.

A better approach is to use optimized transmission conditions, see for example [1, 5]. Using the same configuration as in the previous experiment, and $2\Delta x$ overlap, we obtain with first order transmission conditions and choosing the parameters $p = 1$, $q = 1.75$ (for terminology, see [1]) the result shown in Fig. 5. This illustrates well that using optimized transmission conditions can lead to even better performance of the algorithm than very generous overlap, at no additional cost, since the subdomain size and matrix sparsity is the same as for the case of Dirichlet transmission conditions. In addition we observe that now the convergence has become more linear, and the

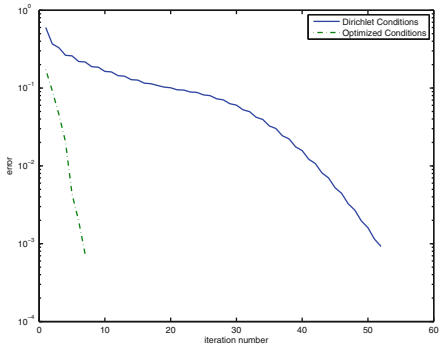


Fig. 5. Comparison of the parareal Schwarz waveform relaxation algorithm with Dirichlet and optimized transmission conditions

algorithm does not depend significantly any more on the superlinear convergence mechanism essential with Dirichlet transmission conditions.

4 Conclusion

We presented a general parareal Schwarz waveform relaxation algorithm, which is based on a decomposition in space and time of a given evolution problem, in order to increase parallelism. We stated a theoretical convergence result, whose proof will appear elsewhere, and then illustrated the dependence of the algorithm on the space-time decomposition configuration, which revealed that for fast convergence, either short time intervals, large overlap, or optimized transmission conditions need to be used. We are currently working on precise convergence factor estimates, a variant of the algorithm which also uses a coarse spatial mesh, and the addition of a coarse propagation mechanism over many spatial subdomains.

Bibliography

[1] Daniel Bennequin, Martin J. Gander, and Laurence Halpern. A homographic best approximation problem with application to optimized Schwarz waveform relaxation. *Math. of Comp.*, 78(265):185–232, 2009.

[2] Philippe Chartier and Bernard Philippe. A parallel shooting technique for solving dissipative ODEs. *Computing*, 51:209–236, 1993.

[3] Ashish Deshpande, Sachit Malhotra, Craig C. Douglas, and Martin H. Schultz. A rigorous analysis of time domain parallelism. *Parallel Algorithms and Applications*, 6:53–62, 1995.

[4] Martin J. Gander and Ernst Hairer. Nonlinear convergence analysis for the parareal algorithm. In U. Langer, M. Discacciati, D.E. Keyes, O.B. Widlund, and W. Zulehner, editors, *Domain Decomposition Methods in Science and Engineering XVII*, volume 60, pages 45–56. Springer-Verlag, 2007.

[5] Martin J. Gander and Laurence Halpern. Optimized Schwarz waveform relaxation methods for advection reaction diffusion problems. *SIAM J. Numer. Anal.*, 45(2):666–697, 2007.

[6] Martin J. Gander, Laurence Halpern, and Frédéric Nataf. Optimal Schwarz waveform relaxation for the one dimensional wave equation. *SIAM J. Numer. Anal.*, 41(5):1643–1681, 2003.

[7] Martin J. Gander, Yao-Lin Jiang, Rong-Jian Li, and Bo Song. A family of parareal Schwarz waveform relaxation algorithms. *In preparation*, 2012.

[8] Martin J. Gander and Andrew M. Stuart. Space-time continuous analysis of waveform relaxation for the heat equation. *SIAM J. Sci. Comput.*, 19(6):2014–2031, 1998.

[9] Martin J. Gander and Stefan Vandewalle. Analysis of the parareal time-parallel time-integration method. *SIAM J. Sci. Comput.*, 29(2):556–578, 2007.

[10] Martin J. Gander and Hongkai Zhao. Overlapping Schwarz waveform relaxation for the heat equation in n-dimensions. *BIT*, 42(4):779–795, 2002.
[11] Wolfgang Hackbusch. Parabolic multi-grid methods. In Roland Glowinski and Jacques-Louis Lions, editors, *Computing Methods in Applied Sciences and Engineering, VI*, pages 189–197. North-Holland, 1984.
[12] Yao-Lin Jiang. A general approach to waveform relaxation solutions of differential-algebraic equations: the continuous-time and discrete-time cases. *IEEE Trans. Circuits and Systems - Part I*, 51(9):1770–1780, 2004.
[13] Yao-Lin Jiang. *Waveform Relaxation Methods*. Scientific Press, Beijing, 2010.
[14] Yao-Lin Jiang and Hui Zhang. Schwarz waveform relaxation methods for parabolic equations in space-frequency domain. *Computers and Mathematics with Applications*, 55(12):2924–2933, 2008.
[15] Ekachai Lelarasmee, Albert E. Ruehli, and Alberto L. Sangiovanni-Vincentelli. The waveform relaxation method for time-domain analysis of large scale integrated circuits. *IEEE Trans. on CAD of IC and Syst.*, 1:131–145, 1982.
[16] Jacques-Louis Lions, Yvon Maday, and Gabriel Turinici. A parareal in time discretization of pde's. *C.R. Acad. Sci. Paris, Serie I*, 332:661–668, 2001.
[17] C. Lubich and A. Ostermann. Multi-grid dynamic iteration for parabolic equations. *BIT*, 27(2):216–234, 1987.
[18] Yvon Maday and Gabriel Turinici. The parareal in time iterative solver: a further direction to parallel implementation. In U. Langer, M. Discacciati, D.E. Keyes, O.B. Widlund, and W. Zulehner, editors, *Domain Decomposition Methods in Science and Engineering XVII*, volume 60, pages 441–448. Springer-Verlag, 2007.
[19] Willard L. Miranker and Werner Liniger. Parallel methods for the numerical integration of ordinary differential equations. *Math. Comp.*, 91:303–320, 1967.
[20] Jörg Nievergelt. Parallel methods for integrating ordinary differential equations. *Comm. ACM*, 7:731–733, 1964.
[21] Stefan Vandewalle and Eric Van de Velde. Space-time concurrent multigrid waveform relaxation. *Ann. Numer. Math.*, 1(1–4):347–363, 1994.
[22] David E. Womble. A time-stepping algorithm for parallel computers. *SIAM J. Sci. Stat. Comput.*, 11(5):824–837, 1990.

A Parallel Overlapping Time-Domain Decomposition Method for ODEs

Stefan Güttel

University of Geneva, Department of Mathematics, 1–2 rue du Lievre, CH-1202 Geneva,
stefan.guettel@unige.ch

Summary. We introduce an overlapping time-domain decomposition for linear initial-value problems which gives rise to an efficient solution method for parallel computers without resorting to the frequency domain. This parallel method exploits the fact that homogeneous initial-value problems can be integrated much faster than inhomogeneous problems by using an efficient Arnoldi approximation for the matrix exponential function.

1 Introduction

We are interested in the parallel solution of a linear initial-value problem

$$u'(t) = Au(t) + g(t), \quad t \in [0,T], \quad u(0) = u_0, \tag{1}$$

where $A \in \mathbb{R}^{N\times N}$ is a possibly large (and sparse) matrix and $u, g : t \mapsto \mathbb{R}^N$. Throughout this paper we assume that the function $g(t)$ is a source term which is difficult to integrate numerically (e.g., highly oscillating or given by a slow computer subroutine). For example, if (1) arises from the space discretization of a heat-diffusion problem, then A represents a diffusion operator and $g(t)$ is a time-dependent heat source.

Problems of the above form arise often in scientific computing, and various solution methods for parallel computers have been proposed in the literature. A popular approach (see, e.g., [1, 8]) is based on the Laplace-transformed equation

$$s\widehat{u}(s) - u_0 = A\widehat{u}(s) + \widehat{g}(s)$$

and the contour integral representation of the inverse transformation

$$u(t) = \frac{1}{2\pi i}\int_\Gamma e^{ts}\widehat{u}(s)\,\mathrm{d}s,$$

with a suitable contour Γ surrounding the singularities of $\widehat{u}(s)$ (which are the eigenvalues of A and all singularities of $\widehat{g}(s)$). Discretization of this integral by a quadrature formula with complex nodes s_j and weights w_j yields

R. Bank et al. (eds.), *Domain Decomposition Methods in Science and Engineering XX*, 459
Lecture Notes in Computational Science and Engineering 91,
DOI 10.1007/978-3-642-35275-1_54,

$$u(t) \approx \sum_{j=1}^{p} w_j \widehat{u}(s_j) = \sum_{j=1}^{p} w_j (s_j I - A)^{-1}(u_0 + \widehat{g}(s_j)).$$

This method is suitable for parallel computation because the p complex shifted linear systems are decoupled. On the other hand, there are obvious drawbacks such as the introduction of complex arithmetic into a real problem and the need for calculating $\widehat{g}(s_j)$. Moreover, many nodes s_j may be required to represent a stiff source $g(t)$ to prescribed accuracy.

Another approach, perhaps closest in spirit to the method described here, is known as exponential quadrature. It is based on the variation-of-constants formula

$$u(t) = e^{tA}u_0 + \int_0^t e^{(t-\tau)A} g(\tau)\,\mathrm{d}\tau$$

and the approximation of the integrand by a quadrature rule in nodes $\tau_1, \ldots, \tau_p$. This yields $p+1$ independent matrix exponentials

$$e^{tA}u_0 \quad \text{and} \quad e^{(t-\tau_j)A} g(\tau_j) \quad \text{for} \quad j = 1, \ldots, p,$$

each of which may be approximated efficiently by a Krylov method (see the discussion in Sect. 3). However, exponential quadrature is impractical if the source term $g(t)$ is stiff enough so that too many quadrature nodes are needed.

To overcome the problems mentioned above, we propose in Sect. 2 a decomposition of (1) into subproblems on overlapping time intervals. These subproblems are decoupled and can be assigned to independent processors. Our method requires almost no communication or synchronization between the processors, except a summation step at the end of the algorithm. Another advantage of our method is its ease of implementation; any available serial integrator for (1) can be used in black-box fashion. Because the efficiency of our method relies on the fast integration of homogeneous linear initial-value problems, Sect. 3 contains a brief discussion of the Arnoldi method for computing the matrix exponential function. In Sect. 4 we discuss the error control and parallel efficiency of our method. In Sect. 5 we present results of a numerical experiment.

2 Overlapping Time-Domain Decomposition

On a time grid $\{T_j = jT/p : j = 0, \ldots, p\}$ we decompose (1) into the following subproblems of two types.

Type 1 : For $j = 1, \ldots, p$ solve

$$v_j'(t) = Av_j(t) + g(t), \quad v_j(T_{j-1}) = 0, \quad t \in [T_{j-1}, T_j],$$

using some serial integrator.

Type 2 : For $j = 1, \ldots, p$ solve

$$w_j'(t) = Aw_j(t), \quad w_j(T_{j-1}) = v_{j-1}(T_{j-1}), \quad t \in [T_{j-1}, T],$$

using exponential propagation (we set $v_0(T_0) := u_0$).

Note that the p subproblems of Type 1 are completely decoupled due to the homogeneous initial values. The same is true for each subproblem of Type 2, the exact solution of which can be computed as

$$w_j(t) = e^{(t-T_{j-1})A} v_{j-1}(T_{j-1}) \tag{2}$$

as soon as the initial value $v_{j-1}(T_{j-1})$ is available. Therefore it is natural to assign the integrations for v_{j-1} and w_j to the same processor so that there is no need for communication and synchronization between the two types of subproblems. Note that the time intervals $[T_{j-1}, T]$ for the w_j are overlapping (see also Fig. 1). By superposition, the solution of (1) is

$$u(t) = v_k(t) + \sum_{j=1}^{k} w_j(t) \quad \text{with } k \text{ such that} \quad t \in [T_{k-1}, T_k].$$

Only the computation of this sum requires communication between the processors. Our parallel algorithm is given by simultaneously integrating the subproblems of Type 1 and Type 2, and finally forming the sum for $u(t)$ at the required time points t.

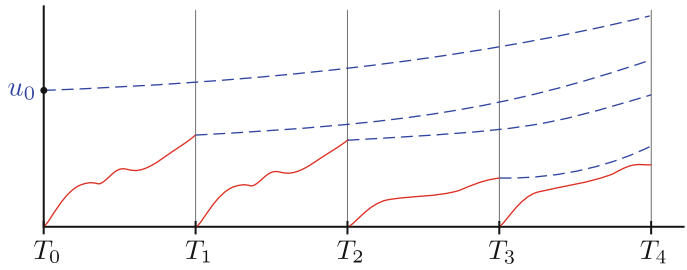


Fig. 1. Time-domain decomposition of an initial-value problem into inhomogeneous subproblems with zero initial value (Type 1, *solid red curves*) and overlapping homogeneous subproblems (Type 2, *dashed blue curves*). The solution is obtained as the sum of all curves

3 Computing the Matrix Exponential

The overlapping propagation of the linear homogeneous subproblems of Type 2 is clearly redundant. To obtain an efficient parallel method, we require that the computation of the matrix exponentials in (2) is fast compared to the integration of the subproblems of Type 1.

For scalar problems ($N = 1$) the computation of the exponential is a trivial task. For computing the exponential of small to medium-sized dense matrices ($N \lessapprox 500$) there are various methods available, see the review [5] and the monograph [4].

The computations become more challenging when the problem size N gets large, in which case the matrix A should be sparse. Then one has to make use of the fact that not the matrix exponential $\exp(tA)$ itself is required, but only the product $\exp(tA)v_0$ with a vector v_0, by using a polynomial or rational Krylov method (see [3] and the references therein). For brevity we will only describe a variant of the restricted-denominator Arnoldi method described in [6] (see also [9]), which extracts an approximation $f_n(t) \approx \exp(tA)v_0$ from a Krylov space built with the matrix $S = (I - A/\sigma)^{-1}A$,

$$\mathscr{K}_n(S, v_0) = \operatorname{span}\{v_0, Sv_0, \ldots, S^{n-1}v_0\},$$

the choice of the parameter $\sigma \in (\mathbb{R} \cup \{\infty\}) \setminus (\Lambda(A) \cup \{0\})$ being dependent on the spectral properties of A. For $\sigma = \infty$ we obtain a standard Krylov space with the matrix A, i.e., $\mathscr{K}_n(S, v_0) = \mathscr{K}_n(A, v_0)$. If $\mathscr{K}_n(S, v_0)$ is of full dimension n, as we assume in the following, we can compute an orthonormal basis $V_n = [v_1, v_2, \ldots, v_n]$ by using the well-known Arnoldi orthogonalization process (see, e.g., [2, Sect. 9.3.5]). The Arnoldi approximation of $\exp(tA)v_0$ is then defined as

$$f_n(t) := V_n \exp(t\,(S_n^{-1} + I_n/\sigma)^{-1})V_n^* v_0, \quad S_n := V_n^* S V_n.$$

Provided that n is small, the computation of $f_n(t)$ requires the evaluation of a $n \times n$ matrix function which is small compared to the original $N \times N$ matrix exponential. Moreover, the matrix S_n can be constructed without explicit projection from quantities computed in the Arnoldi process.

In Fig. 2 we show the error norm $\|\exp(A)v_0 - f_n(1)\|_2$ of the Arnoldi approximations with parameters $\sigma = \infty$ and $\sigma = 40$ (a rather arbitrary choice) as a function of n, for the matrices

$$A_1 = \operatorname{tridiag}(30, -40, 10) \in \mathbb{R}^{199\times 199},\ A_2 = \operatorname{tridiag}(60, -90, 30) \in \mathbb{R}^{299\times 299}$$

arising from the finite-difference discretization of the same 1D advection–diffusion problem, and a random vector v_0. We have also plotted the error of orthogonal projection of the exact solution onto the space $\mathscr{K}_n(S, v_0)$, namely $V_n V_n^* e^A v_0$, and observe that the Arnoldi method is capable of extracting an approximation nearby this projection. For comparison we show the error of the result produced by n steps of various explicit and implicit integrators for the initial-value problem $v' = Av$, $v(0) = v_0$, integrated to $t = 1$. For this linear homogeneous problem all integrators actually compute approximations from some Krylov space $\mathscr{K}_n(S, v_0)$ (for the explicit integrators with shift $\sigma = \infty$ and for implicit Euler with $\sigma = n$), but the Arnoldi methods extract much better approximations in the same number of iterations. Note also that the Arnoldi method with finite shift $\sigma = 40$ converges almost independently of the problem size N, a property often referred to as *mesh-independence*.

Because the error of Arnoldi approximations decays usually very fast (i.e., $\|e^{tA}v_0 - f_{n+1}(t)\|$ is considerably smaller than $\|e^{tA}v_0 - f_n(t)\|$), it is often sufficient

to use the difference of two consecutive iterates as an estimate for the approximation error:

$$\begin{aligned} \|e^{tA}v_0 - f_n(t)\| &\le \|e^{tA}v_0 - f_{n+1}(t)\| + \|f_{n+1}(t) - f_n(t)\| \\ &\approx \|f_{n+1}(t) - f_n(t)\|. \end{aligned} \tag{3}$$

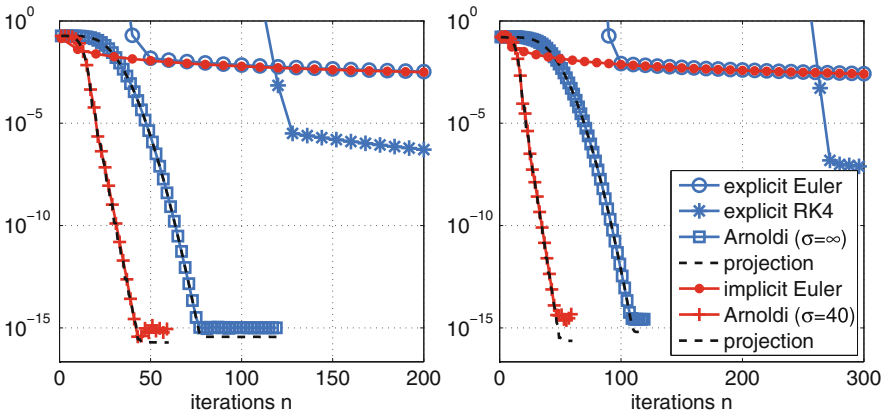


Fig. 2. Error (2-norm) of various time-stepping methods and Krylov methods for a linear homogeneous advection–diffusion problem $v' = Av$, $v(0) = v_0$, of size $N = 199$ (*left*) and $N = 299$ (*right*) as a function of time steps or Krylov space dimension n, respectively

4 Error Control and Parallel Efficiency

Many ODE solvers, for example those of MATLAB, use an error control criterion like

$$\|e(t)\|_\infty \le \max\{\texttt{reltol} \cdot \|\widetilde{u}(t)\|_\infty, \texttt{abstol}\}, \quad t \in [0,T],$$

where $e(t) = u(t) - \widetilde{u}(t)$ is the (estimated) error of the computed solution $\widetilde{u}(t)$. Because the inhomogeneous subproblems of Type 1 for $v_j(t)$ are solved with zero initial guess, it is not advisable to use an error criterion which is relative to the norm of the solution. Hence we assume that all of these subproblems are solved with an absolute error $\|e_j(t)\|_\infty \le \texttt{abstol}/p$ over the time interval $[T_{j-1}, T_j]$. This error is then propagated exponentially over the remaining interval $[T_j, T]$, hence we have to study the transient behavior of

$$\|e^{tA}e_j(T_j)\|_\infty \le \|e^{tA}\|_\infty \texttt{abstol}/p \tag{4}$$

for $t \in [0, T - T_j]$. It is well known that for a *stable* matrix A (i.e., all eigenvalues lie in the left complex half-plane) the limit $\lim_{t\to\infty} \|e^{tA}\|_\infty$ is finite. Unfortunately, the

norm may initially grow arbitrarily large before convergence sets in, a phenomenon usually referred to as *hump* (see [5]). However, for a diagonally dominant matrix $A = (a_{ij})$ with $a_{ii} \le 0$ this cannot happen, as one can show as follows (cf. [7]): Define $\rho = \max_i\{a_{ii} + \sum_{j\neq i}|a_{ij}|\} \le 0$. By the formula $\exp(tA) = \lim_{k\to\infty}(I + tA/k)^k$ we have $\|e^{tA}\|_\infty \le \lim_{k\to\infty}\|I + tA/k\|_\infty^k$. For k sufficiently large we have

$$\|I + tA/k\|_\infty = \max_i \left\{1 + t\left(a_{ii} + \sum_{j\neq i}|a_{ij}|\right)/k\right\} = 1 + t\rho/k,$$

hence

$$\|e^{tA}\|_\infty \le \lim_{k\to\infty}(1 + t\rho/k)^k = e^{t\rho} \le 1 \quad \text{for all } t \ge 0.$$

Of course, it is possible to estimate the behavior of $\|e^{tA}\|$ for general matrices and in other norms (see, e.g., [10]), but for brevity we will only consider a diagonally dominant A. In this case the errors $e_j(t)$ of the subproblem solutions $v_j(t)$ $(j = 1,\dots,p)$ are non-increasing when being exponentially propagated, and if we assume that the subproblems of Type 2 are solved exactly (or with sufficiently high accuracy), then the overall error $e(t)$ is bounded[1] by the sum of subproblem errors (4), hence $\|e(t)\|_\infty \le$ `abstol`. If the integrator is a time-stepping method of order q, it is reasonable to assume that the computation time for one subproblem of Type 1 is at most $\tau_1(p) = (\tau_0 \cdot p^{1/q})/p$, where τ_0 is the computation time for serial integration over $[0,T]$. If each subproblem of Type 2 takes at most τ_2 units of computation time, the expected efficiency of our parallel algorithm is at least

$$\text{efficiency} = \frac{\text{speedup}}{p} = \frac{1}{p}\cdot\frac{\tau_0}{\tau_1(p)+\tau_2} = \left(p^{1/q} + \frac{p\cdot\tau_2}{\tau_0}\right)^{-1}. \tag{5}$$

The efficiency becomes large if the serial computation time τ_0 is long compared to $p\cdot\tau_2$, and if the integration order q is high.

5 Numerical Example

As a simple model problem we consider the 1D heat equation

$$\begin{aligned}
\partial_t u(t,x) &= \alpha\,\partial_{xx}u(t,x) + g(t,x) && \text{on } x \in (0,1),\\
u(t,0) &= u(t,1) = 0,\\
u(0,x) &= u_0(x) = 4x(1-x),\\
g(t,x) &= e\max\{1 - |c-x|/d, 0\}, && \text{where } c = .5 + (.5-d)\sin(2\pi f t).
\end{aligned}$$

The source term $g(t,x)$ is a hat function centered at c with half-width $d = 0.05$ and height $e = 100\cdot\alpha^{1/2}$, oscillating with frequency f. Finite-difference discretization

[1] This worst-case bound is sharp only if all errors e_j are collinear, which is rather unlikely. Probabilistic error estimation would give $\|e(t)\|_\infty \lessapprox$ `abstol`$/\sqrt{p}$. This explains why the observed parallel efficiency of our algorithm is usually better than predicted by (5). We plan to investigate this in a sequel.

at $N = 100$ points $x_j = j/(N+1)$ $(j = 1,\ldots,N)$ yields an initial-value problem (1), where $A = \alpha(N+1)^2\,\mathrm{tridiag}(1,-2,1) \in \mathbb{R}^{N\times N}$. This problem is integrated over the time interval $[0, T = 1]$. For the serial integration we have used the classical Runge–Kutta method of order $q = 4$ (implemented in MATLAB) with constant step size

$$h_0 = \min\{5\cdot 10^{-5}/\alpha, 10^{-2}/f\},$$

chosen to avoid instability of the time-stepping method caused by the stiff linear term $Au(t)$ and to capture the oscillations of $g(t)$. As shown in Table 1, the absolute error (∞-norm) is at most $5\cdot 10^{-4}$ for all diffusion coefficients $\alpha = 0.01, 0.1, 1$ and frequencies $f = 1, 10, 100$. These parameters determine the stiffness of $Au(t)$ and $g(t)$, respectively. We have also tabulated the serial integration times τ_0. As expected, these are roughly proportional to h_0^{-1}.

For our parallel algorithm we have partitioned the interval $[0, T]$ in $p = 4$ subintervals, and computed the solution $u(t)$ at all time points $T_j = jT/p$ $(j = 1,\ldots,p)$. The subproblems of Type 1 are integrated with step size $h_1 = h_0/\sqrt{p}^{1/q}$ (based on a probabilistic error assumption, see the footnote on p. 464). In Table 1 we list the maximal computation time τ_1 for all subproblems of Type 1 among all processors.

For the subproblems of Type 2 we have used the Arnoldi method described in Sect. 3 with shift $\sigma = 5.3$, in combination with the ∞-norm error estimate (3) for an accuracy of 10^{-4} (for more details on the selection of σ we refer to [9]). In Table 1 we list the maximal computation time τ_2 for all subproblems of Type 2 among all processors.

The errors of the final solutions computed with our parallel algorithm are shown in the second-last column, and they are all below the errors obtained by sequential integration. This indicates that our choice for the step size h_1 is reasonable. The parallel efficiency of our algorithm is above 50 % for all nine tests, and it increases with frequency f because smaller time steps are required to integrate the inhomogeneity accurately. We finally note that for large-scale computations our algorithm could also be used to further speed up a saturated space parallelization (e.g., by domain decomposition).

Acknowledgments I am grateful to Martin J. Gander for many helpful discussions and valuable comments.

Bibliography

[1] I. P. Gavrilyuk and V. L. Makarov. Exponentially convergent algorithms for the operator exponential with applications to inhomogeneous problems in Banach spaces. *SIAM J. Numer. Anal.*, 43:2144–2171, 2005.

[2] G. H. Golub and C. F. Van Loan. *Matrix Computations*. Johns Hopkins University Press, Baltimore, MD, 3rd edition, 1996.

[3] S. Güttel. *Rational Krylov Methods for Operator Functions*. PhD thesis, Institut für Numerische Mathematik und Optimierung der Technischen Universität Bergakademie Freiberg, 2010.

Table 1. Serial and parallel performance with $p = 4$ processors for a heat equation with diffusion coefficient α and source-term frequency f.

α	f	serial		parallel			efficiency
		τ_0	error	τ_1	τ_2	error	
0.01	1	4.97e−02	3.01e−04	1.58e−02	9.30e−03	2.17e−04	50 %
0.01	10	2.43e−01	4.14e−04	7.27e−02	9.28e−03	1.94e−04	74 %
0.01	100	2.43e+00	1.73e−04	7.19e−01	9.26e−03	5.68e−05	83 %
0.1	1	4.85e−01	2.24e−05	1.45e−01	9.31e−03	5.34e−06	79 %
0.1	10	4.86e−01	1.03e−04	1.45e−01	9.32e−03	9.68e−05	79 %
0.1	100	2.42e+00	1.29e−04	7.21e−01	9.24e−03	7.66e−05	83 %
1	1	4.86e+00	7.65e−08	1.45e+00	9.34e−03	1.78e−08	83 %
1	10	4.85e+00	8.15e−06	1.45e+00	9.33e−03	5.40e−07	83 %
1	100	4.85e+00	3.26e−05	1.44e+00	9.34e−03	2.02e−05	84 %

[4] N. J. Higham. *Functions of Matrices. Theory and Computation*. SIAM, Philadelphia, PA, 2008.

[5] C. Moler and C. F. Van Loan. Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later. *SIAM Rev.*, 45:3–39, 2003.

[6] I. Moret and P. Novati. RD-rational approximations of the matrix exponential. *BIT*, 44:595–615, 2004.

[7] D. L. Powers and R. Jeltsch. Problem 74–5: On the norm of a matrix exponential. *SIAM Rev.*, 17:174–176, 1975.

[8] D. Sheen, I. H. Sloan, and V. Thomée. A parallel method for time discretization of parabolic equations based on Laplace transformation and quadrature. *IMA Journal of Numerical Analysis*, 23:269–299, 2003.

[9] J. van den Eshof and M. Hochbruck. Preconditioning Lanczos approximations to the matrix exponential. *SIAM J. Sci. Comput.*, 27:1438–1457, 2006.

[10] C. F. Van Loan. The sensitivity of the matrix exponential. *SIAM J. Numer. Anal.*, 14:971–981, 1977.

Two-Grid LNKSz for Distributed Control of Unsteady Incompressible Flows

Haijian Yang[1] and Xiao-Chuan Cai[2]

[1] College of Mathematics and Econometrics, Hunan University, Changsha, Hunan 410082, P. R. China, haijianyang@gmail.com
[2] Department of Computer Science, University of Colorado at Boulder, Boulder, CO 80309, USA, cai@cs.colorado.edu

Summary. The distributed control of unsteady incompressible flows has been the focus of intense research in scientific computing in the past few years. Most of the existing approaches for distributed control problems are based on the so-called reduced space method which is easier to implement but may have convergence issues in some situations. In this paper we investigate some fully coupled parallel two-grid Lagrange-Newton-Krylov-Schwarz (LNKSz) algorithms for the implicit solution of distributed control problems. In the full space approach we couple the control variables, the state variables and the adjoint variables in a single large system of nonlinear equations. Numerical experiments are presented to show the efficiency and scalability of the algorithm on supercomputers with more than one thousand processors.

1 Introduction

Flow optimal control problems have many important applications in science and engineering and many attempts have been made in the past few years to mathematically understand and numerically solve flow control problems in various forms; see e.g., [3, 6]. Popular approaches for solving unsteady flow control problems are explicit or semi-implicit methods, both are limited by a Courant-Friedrichs-Lewy (CFL) condition. Recently, the class of full space Lagrange-Newton-Krylov-Schwarz (LNKSz) algorithms was introduced for solving the steady state flow control problem [4, 5]. The methods include two parts: a Lagrange-Newton method for the nonlinear system obtained from the optimization problem and a Krylov subspace method for the Jacobian system arising from the Newton method. In this paper we propose a class of fully coupled parallel two-grid Lagrange-Newton-Krylov-Schwarz (LNKSz) algorithms for the distributed control of unsteady incompressible flows. Since we use a fully implicit scheme, the CFL condition can be completely relaxed. We show numerically that the proposed LNKSz is stable and converges well with relatively large times steps, and it is robust with respect to some of the physical parameters, such as the Reynolds number.

The rest of the paper is organized as follows. In Sect. 2, we present the unsteady distributed control problems and introduce a fully implicit discretization scheme.

R. Bank et al. (eds.), *Domain Decomposition Methods in Science and Engineering XX*, Lecture Notes in Computational Science and Engineering 91,
DOI 10.1007/978-3-642-35275-1_55,

Section 3 includes the main components and features of LNKSz. Some numerical results are given in Sect. 4. We end the paper with some concluding remarks in Sect. 5.

2 Mathematical Model and Discretization

We consider the two-dimensional unsteady incompressible Navier-Stokes equations in the velocity-vorticity formulation:

$$\begin{cases} -\Delta v_1 - \dfrac{\partial \omega}{\partial y} = 0 \ \text{ in } [0,T]\times\Omega, \\ -\Delta v_2 + \dfrac{\partial \omega}{\partial x} = 0 \ \text{ in } [0,T]\times\Omega, \\ \dfrac{\partial \omega}{\partial t} - \dfrac{1}{Re}\Delta\omega + v_1\dfrac{\partial \omega}{\partial x} + v_2\dfrac{\partial \omega}{\partial y} - \operatorname{curl}\mathbf{f} = 0 \ \text{ in } [0,T]\times\Omega, \end{cases} \tag{1}$$

where Ω is the computational domain and $[0,T]$ is the time interval. In the above equations the velocity field $\mathbf{v} = (v_1, v_2)$ and the vorticity ω are the state variables, $\mathbf{f} = (f_1, f_2)$ is the external force, $\operatorname{curl}\mathbf{f} = -\partial f_1/\partial y + \partial f_2/\partial x$, and Re is the Reynolds number.

In the distributed control problem we try to find an external force $\mathbf{f}$ over the control domain $\Omega_f \subseteq \Omega$ in order to achieve the goal

$$\min \mathscr{F}(\mathbf{v},\omega,\mathbf{f}) = \frac{1}{2}\int_0^T \mathscr{G}(\mathbf{v},\omega)\,dt + \frac{\gamma}{2}\int_0^T \int_{\Omega_f} \|\mathbf{f}\|_2^2 \, d\Omega \, dt \tag{2}$$

subject to the constraints (1) with some initial and boundary conditions. Here, $\mathscr{G}(\mathbf{v},\omega)$ is the objective function of the optimal control problem, $\gamma > 0$ is a regularization parameter used to restrict the magnitude of the external force so that it is not unrealistically large.

For solving unsteady distributed control problems, it typically requires a combination of a discretization in space and time with an optimization method. In this paper we follow the discretize-then-optimize approach with a finite difference method for the space discretization and a second-order backward differentiation formula for the time discretization. The original full-time-interval problem is too expensive to solve even on the latest supercomputers, we therefore replace it by a sequence of suboptimal problems, which are similar to the original problem but only defined on the time interval $[t^{(k-1)}, t^{(k)}]$, $k = 1, 2, \ldots, k_{\max}$, with $t^{(0)} = 0$ and $t^{(k_{\max})} = T$. Let $\mathbf{x} = (\mathbf{v}, \omega, \mathbf{f})$. Then on each time interval we write the discrete suboptimzation problem as follows:

$$\begin{cases} \min \ \mathscr{F}_h^{(k)}(\mathbf{x}) \\ \text{s.t. } \mathbf{C}_h^{(k)}(\mathbf{x}) = \mathbf{0}, \end{cases} \tag{3}$$

where $\mathscr{F}_h^{(k)}(\mathbf{x})$ is the restriction of $\mathscr{F}$ on the interval $[t^{(k-1)}, t^{(k)}]$, and $\mathbf{C}_h^{(k)}(\mathbf{x})$ are the constraints defined on the time interval $[t^{(k-1)}, t^{(k)}]$.

By introducing the Lagrange multipliers λ with respect to the state and control variables, we define the following Lagrangian functional

$$\mathscr{L}^{(k)}(\mathbf{x},\lambda) \equiv \mathscr{F}_h^{(k)}(\mathbf{x}) + \left(\lambda, \mathbf{C}_h^{(k)}(\mathbf{x})\right). \tag{4}$$

Let $X \equiv (\mathbf{x},\lambda)$. Then, for $k = 1,2,\ldots,k_{\max}$, the KKT system obtained by differentiating (4) becomes

$$G^{(k)}(X) = \begin{pmatrix} \nabla_{\mathbf{x}} \mathscr{L}^{(k)}(\mathbf{x},\lambda) \\ \nabla_{\lambda} \mathscr{L}^{(k)}(\mathbf{x},\lambda) \end{pmatrix} = 0. \tag{5}$$

The optimality system (5) is a large, nonlinear, coupled, and muti-components system. Moreover, the corresponding Jacobian matrix is indefinite and very ill-conditioned. Hence, a good preconditioner is essential to solve the optimality system efficiently.

3 Two-Grid Newton Method and Schwarz Preconditioners

The class of full space LNKSz method includes the following steps: the Lagrangian functional is formed and differentiated to obtain the KKT system; then the inexact Newton method with line search is applied; and at each Newton iteration the linear system is solved with a one-level or two-level Schwarz preconditioned Krylov subspace method. We refer to LNKSz combined with the one-level (two-level) Schwarz preconditioner as one-level (two-level) LNKSz method.

When using Newton's method to solve the nonlinear system (5) on a grid, one of the major problems is the deterioration of the convergence rate when the grid is refined, specially for the first time step, since in this case the initial guess is not good enough for the Newton iterations. After many experiments, we find that a solution to the problem is "grid-sequencing", which is quite effective in keeping the number of nonlinear iterations small. In order to use grid-sequencing, we assume there are two grids covering Ω, a coarse grid of size H and a fine grid of size h. We first use the one-level method to solve the nonlinear problem on the coarse grid with the initial guess obtained as a restriction of the fine grid solution from the previous timestep. Of course, at the first time step, we choose the initial condition as the initial guess. Then, we interpolate the solution to the fine grid and use it as an initial guess for the nonlinear problem on the fine grid. We refer to this LNKSz method combined with the grid-sequencing technique as the two-grid LNKSz method in which the same coarse grid is also used to build the two-level Schwarz preconditioner for solving the Jacobian problem.

We assume that Ω is covered by a non-overlapping and an overlapping partition as in [2]. Let J be the Jacobian matrix of the nonlinear problem (5) on the fine grid and let R_i^{δ} and R_i^0 be the restriction operator from Ω to its overlapping and non-overlapping subdomains, respectively. Here δ is the size of the overlap. Then the one-level restricted additive Schwarz (RAS) preconditioner [2] is defined as

$$M_{RAS}^{-1} = \sum_{i=1}^{N_p} (R_i^0)^T J_i^{-1} R_i^\delta. \tag{6}$$

with $J_i = R_i^\delta \; J \; (R_i^\delta)^T$ and N_p is the number of subdomains, which is the same as the number of processors. Let J_c be the Jacobian matrix on the coarse grid and I_h^H a restriction operator from the fine grid to the coarse grid. Then a multiplicative type two-level Schwarz preconditioner [8, 9] is defined as

$$M^{-1} = \Big(I - (I - M_{RAS}^{-1}J)(I - M_c^{-1}J)(I - M_{RAS}^{-1}J)\Big)J^{-1} \tag{7}$$

with $M_c^{-1} = (I_h^H)^T J_c^{-1} I_h^H$ and I is the identity matrix.

4 Numerical Experiments

Our algorithms are implemented based on the Portable Extensible Toolkit for Scientific computing (PETSc) [1]. All computations are performed on an IBM BlueGene/L supercomputer.

In the following, we describe a backward-facing step flow control problem [7]. Let $\Omega = (0,6) \times (0,1)$, $\Omega_f = (0,1) \times (0,0.5)$, $T = 1$, Γ be the boundary of the domain Ω, $\Gamma_2 = \{(x,y) \in \Gamma : 0 < y < 1, x = 6\}$, $\Gamma_4 = \{(x,y) \in \Gamma : 0 < y < 1, x = 0\}$, and $\Gamma_{4,a} = \{(x,y) \in \Gamma_4 : 0.5 \le y < 1\}$. Then the backward-facing step control problem consists of finding $(v_1, v_2, \omega, f_1, f_2)$ such that the minimization

$$\min \mathscr{F}(\omega, \mathbf{f}) = \frac{1}{2}\int_0^T \int_\Omega \omega^2 \, d\Omega \, dt + \frac{\gamma}{2}\int_0^T \int_{\Omega_f} \|\mathbf{f}\|_2^2 \, d\Omega \, dt \tag{8}$$

is achieved subject to the constraints (1) with the following boundary conditions:

$$\begin{cases} v_1 & = v_{in} \quad \text{on } [0,T] \times \Gamma_{4,a}, \\ v_1 & = v_{out} \quad \text{on } [0,T] \times \Gamma_2, \\ v_1 & = 0 \quad \text{on } [0,T] \times \Gamma_u, \\ v_2 & = 0 \quad \text{on } [0,T] \times \Gamma, \\ \omega + \dfrac{\partial v_1}{\partial y} - \dfrac{\partial v_2}{\partial x} & = 0 \quad \text{on } [0,T] \times \Gamma, \\ \mathbf{v}(0,x,y) - \mathbf{v}_0 & = \mathbf{0} \quad \text{in } \overline{\Omega}, \\ \omega(0,x,y) + \dfrac{\partial v_{0,1}}{\partial y} - \dfrac{\partial v_{0,2}}{\partial x} & = 0 \quad \text{in } \overline{\Omega}, \end{cases} \tag{9}$$

where $\Gamma_u = \Gamma \backslash (\Gamma_{4,a} \cup \Gamma_2)$. At the inflow boundary, a parabolic velocity profile $v_{in} = 8(1-y)(y-\frac{1}{2})cos(t)$ is imposed. At the outflow boundary, $v_{out} = y(1-y)cos(t)$ is applied. The following initial velocity is defined by $\mathbf{v}_0 = (v_{0,1}, v_{0,2})$ with

$$v_{0,1} = \begin{cases} y(1-y) + \dfrac{1}{16}y & \text{if } 0 \le y \le \dfrac{1}{2}, \\ \\ y(1-y) + \dfrac{1}{16}(1-y) & \text{if } \dfrac{1}{2} \le y \le 1, \end{cases}$$

and $v_{0,2}(x,y) = 0$. The parameter $\gamma = 0.1$.

In the experiments, we compare the following algorithms which are introduced in Sect. 3:

- One-level LNKSz: one-level additive Schwarz is used as the Jacobian solve, and inexact Newton is carried out on the fine grid;
- Two-level LNKSz: two-level multiplicative Schwarz is used as the Jacobian solve, and inexact Newton is carried out on the fine grid;
- Two-grid LNKSz: two-level multiplicative Schwarz is used as the Jacobian solve on the fine grid, inexact Newton is used on the coarse grid to generate the initial guess for the inexact Newton on the fine grid.

In all the experiments, all Jacobian matrices are constructed approximately using a multi-colored finite difference method. The size of the coarse grid H is taken as $4h$, where h is the size of the fine grid. GMRES(90) and FGMRES(90) are used to solve the linear system at each Newton step on the coarse and the fine grids, respectively. In the one-level method, the overlapping size is $\delta = 6$. In the two-level and two-grid methods, the overlapping sizes of the coarse grid and the fine grids are $\delta_c = 4$ and $\delta = 6$, respectively. There are several nested iterative procedures in the proposed algorithms, and each requires a proper stopping condition. We use 10^{-10} (10^{-6}) as the absolute (relative) condition for all linear and nonlinear solves, except for the linear coarse solve of the two-level preconditioner, for which we use 10^{-4} (10^{-2}) as the absolute (relative) condition. The subdomain problems are solved with a sparse LU factorization.

Next, we present results for the test problem and discuss some details of the two-grid LNKSz. First, we compare the three methods in Table 1. Note that, the one-level method doesn't converge when $N_p = 1{,}024$, which is caused by the divergence of GMRES. Moreover, we note that: (1) for the linear solver, the number of GMRES iterations for the one-level LNKSz is much larger than that for the two-level and two-grid methods; (2) for the nonlinear solver, the numbers of Newton iterations for the one-level and two-level methods are also larger than that for the two-grid method; and (3) compared with the one-level and two-level methods, the total computing time for the two-grid method is much smaller. When the Reynolds number increases from 200 to 400, for one-level and two-level methods, the average number of Newton iterations and the total computing time become larger. With the help of grid-sequencing, the convergence of the two-grid method is less sensitive to the Reynolds number. Based on the results of Table 1, it is clear that the two-grid method is better than the others.

An important implementation detail to consider in designing two-grid LNKSz is to balance the quality of the initial guess for the fine grid Newton iterations and the computing time on the coarse solver. In Table 2, we present a comparison of the computing time for the two-level and two-grid methods. In this table, we report the total time spent on the Newton iterations at some time steps, the time spent on the Newton iterations on the coarse solver, and the percentage between these two computational costs. We observe that the cost of Newton iterations on the coarse grid is very small compared with the total computational cost. It is important to note that the coarse

Table 1. A comparison of three methods. 768×128 grid, and $\Delta t = 0.1$ (i.e., there are 10 time steps). "N_p" stands for the number of processors which is the same as the number of subdomains, "IN" is the average number of inexact Newton iterations per time step on the fine grid, "RAS" is the average number of RAS preconditioned GMRES iterations per Newton iteration, and "Time" is the total computing time in seconds. "$**$" means the divergence of GMRES.

N_p	Method	IN	RAS	Time	IN	RAS	Time
		Re=200			Re=400		
64	One-level	3.2	165.4	1370.4	3.7	158.9	1557.5
64	Two-level	3.2	20.4	1342.8	3.7	19.2	1528.0
64	Two-grid	2.1	18.7	898.2	2.0	18.0	836.4
256	One-level	3.2	531.3	795.5	3.7	632.9	1052.3
256	Two-level	3.2	27.4	479.9	3.7	27.1	560.1
256	Two-grid	2.1	25.5	317.5	2.0	26.1	313.2
1024	One-level		**			**	
1024	Two-level	3.2	66.3	314.3	3.7	67.9	376.9
1024	Two-grid	2.1	64.2	208.5	2.0	68.5	209.8

grid has to be sufficiently fine so that the coarse solution has a reasonable accuracy, otherwise, it won't be able to provide a good initial guess for the fine grid nonlinear solver.

Table 2. A comparison of the computing time for the test problem at several different time steps. $Re = 400$, 768×128 grid, and $\Delta t = 0.1$ (i.e., there are 10 time steps). The heading "Timestep(k)" represents the time step k, "Time" is the total time spent on the Newton iterations at the time step k, "Coarse_time" is the time spent on the Newton iterations on the coarse solver at the time step k, and "Percent(%)" is ("Coarse_time"/"Time").

N_p	Timestep(k)	Time	Coarse_time	Percent(%)	Time
			Two-grid		Two-level
64	$k = 1$	110.0	3.87	3.52%	458.9
64	$k = 2$	80.0	2.39	2.99%	117.0
64	$k = 5$	82.5	2.50	3.03%	118.0
64	$k = 10$	84.7	2.51	2.96%	119.0
256	$k = 1$	38.6	1.71	4.43%	172.8
256	$k = 2$	29.7	0.99	3.33%	41.4
256	$k = 5$	30.0	1.04	3.43%	41.6
256	$k = 10$	30.8	1.06	3.44%	42.3
1024	$k = 1$	23.3	1.37	5.88%	115.1
1024	$k = 2$	20.6	0.68	3.30%	28.1
1024	$k = 5$	21.2	0.72	3.39%	28.4
1024	$k = 10$	21.5	0.74	3.44%	30.8

One of the difficulties in the nonlinear solver is the choice of the initial guess. In Fig. 1, we show the nonlinear residual history by using three different methods at the first time step (i.e., $k = 1$). One can see that the nonlinear system is difficult to solve by using one-level or two-level method. In fact, it takes 11 iterations for the one-level or two-level method to converge. By using the two-grid method only three Newton iterations are required to satisfy the desired stopping condition.

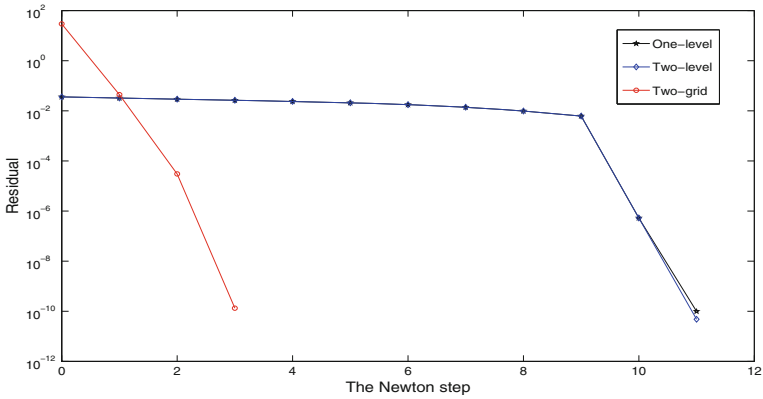


Fig. 1. Nonlinear residual history by using three different methods at the first time step, for $Re = 200$, 768×128 grid and 64 processors, and $\Delta t = 0.1$

5 Conclusions

In this paper, we developed a family of two-grid algorithms for distributed control of unsteady incompressible flows. With the help of the two-grid Newton method and the two-level Schwarz preconditioner, we showed numerically that these strategies provide substantial improvement of the overall method in terms of the total computing time, the number of linear iterations, and the number of Newton iterations, especially when the number of processors is large.

Bibliography

[1] S. Balay, K. Buschelman, W. D. Gropp, D. Kaushik, M. Knepley, L. C. McInnes, B. F. Smith, and H. Zhang. *PETSc Users Manual*. Argonne National Laboratory, 2010.

[2] X.-C. Cai and M. Sarkis. A restricted additive Schwarz preconditioner for general sparse linear systems. *SIAM J. Sci. Comput.*, 21:792–797, 1999.

[3] M. Gunzburger. *Perspectives in Flow Control and Optimization*. SIAM, First ed., 2003.
[4] E. Prudencio and X.-C. Cai. Parallel multilevel restricted Schwarz preconditioners with pollution removing for PDE-constrained optimization. *SIAM J. Sci. Comput.*, 29:964–985, 2007.
[5] E. Prudencio, R. Byrd, and X.-C. Cai. Parallel full space SQP Lagrange-Newton-Krylov-Schwarz algorithms for PDE-constrained optimization problems. *SIAM J. Sci. Comp.*, 27:1305–1328, 2006.
[6] L. Quartapelle. *Numerical Solution of the Incompressible Navier-Stokes Equations*. International Series of Numerical Mathematics, Birkhäuser Verlag, 1996.
[7] S. S. Ravindran. Numerical approximation of optimal control of unsteady flows using SQP and time decomposition. *Int. J. Numer. Meth. Fluids.*, 45:21–42, 2004.
[8] B. Smith, P. Bjørstad, and W. Gropp. *Domain Decomposition: Parallel Multilevel Methods for Elliptic Partial Differential Equations*. Cambridge University Press, New York, NY, 1996.
[9] A. Toselli and O. Widlund. *Domain Decomposition Methods-Algorithms and Theory*. Springer-Verlag, Berlin, 2005.

On the Applicability of Lions' Energy Estimates in the Analysis of Discrete Optimized Schwarz Methods with Cross Points

Martin J. Gander and Felix Kwok

Section de Mathématiques, Université de Genève
{Martin.Gander|Felix.Kwok}@unige.ch

1 Introduction

For a bounded open subset $\Omega \subset \mathbb{R}^2$, suppose we want to solve

$$(\eta - \Delta)u = f \qquad \text{on } \Omega, \qquad u = g \qquad \text{on } \partial\Omega, \tag{1}$$

for $\eta \geq 0$ using the optimized Schwarz method (OSM)

$$\begin{aligned} &(\eta - \Delta)u_i^k = f|_{\Omega_i} \quad \text{on } \Omega_i, \qquad u_i^k = g|_{\partial\Omega_i} \quad \text{on } \partial\Omega_i \cap \partial\Omega, \\ &\frac{\partial u_i^k}{\partial n_i} + p_{ij}u_i^k = \frac{\partial u_j^{k-1}}{\partial n_i} + p_{ij}u_j^{k-1} \qquad \text{on } \Gamma_{ij} \text{ for all } \Gamma_{ij} \neq \emptyset, \end{aligned} \tag{2}$$

for $k = 1, 2, \ldots$ and $i = 1, \ldots, n$, where $\Omega_i \subset \Omega$ are non-overlapping subdomains, $\Gamma_{ij} = \partial\Omega_i \cap \overline{\Omega_j}$ is the interface between Ω_i and an adjacent subdomain Ω_j, $j \neq i$, and $p_{ij} > 0$ are Robin parameters along Γ_{ij}. In [7], the powerful technique of energy estimates is used to show convergence of (2) for $\eta = 0$ under very general conditions. Similar techniques have been used to prove convergence results for other types of equations, cf. [2] for the Helmholtz equation and [5] for the time-dependent wave equation. While one often assumes that the proof carries over trivially to finite-element discretizations, it has been reported in the literature (cf. [8, 9]) that discrete OSMs can diverge when the domain decomposition contains cross points, i.e., when more than two subdomains share a common point. This is in apparent contradiction to Lions' proof, and such difficulties contribute to the limited use of OSMs in practice. The goal of this paper is to explain why the presence of cross points makes it possible for the discrete OSM to diverge despite the proof of convergence at the continuous level, and why this difference in behavior is generally unavoidable.

The remainder of the paper proceeds as follows. In Sect. 2, we recall Lions' energy estimate argument. In Sect. 3, we explain why it is impossible to convert the continuous energy estimate into a discrete one in a generic way, without sacrificing continuity of the solutions across subdomain boundaries. In Sect. 4, we show two modifications that preserve continuity of the discrete solutions, but both must be used

R. Bank et al. (eds.), *Domain Decomposition Methods in Science and Engineering XX*, 475
Lecture Notes in Computational Science and Engineering 91,
DOI 10.1007/978-3-642-35275-1_56,

with Krylov methods to avoid divergent iterations. Finally, we show in Sect. 5 that a Lions-type discrete estimate can only hold under very stringent conditions; thus, continuous estimates generally do not predict the behavior of discrete OSMs.

2 Continuous Energy Estimates

We briefly recall the argument in [7] proving the convergence of (2). We assume $p_{ij} = p_{ji}$ to be a positive function that is bounded away from zero and defined on $\Gamma_{ij} = \Gamma_{ji}$. To show that (2) converges for all initial guesses, we first write the error equations

$$\begin{aligned} &(\eta - \Delta)e_i^k = 0 \quad \text{on } \Omega_i, \qquad e_i^k = 0 \quad \text{on } \partial\Omega \cap \partial\Omega_i, \\ &\frac{\partial e_i^k}{\partial n_i} + p_{ij}e_i^k = \frac{\partial e_j^{k-1}}{\partial n_i} + p_{ij}e_j^{k-1} \quad \text{on } \Gamma_{ij} \text{ for all } \Gamma_{ij} \neq \emptyset, \end{aligned} \tag{3}$$

where $e_i = u_i^k - u|_{\Omega_i}$ with u being the exact solution to (1). We then multiply the first equation in (3) by e_i^k and integrate to get

$$0 = a_i(e_i^k, e_i^k) - \int_{\partial\Omega_i} e_i^k \frac{\partial e_i^k}{\partial n_i} = a_i(e_i^k, e_i^k) - \sum_{(i,j)\in E} \int_{\Gamma_{ij}} e_i^k \frac{\partial e_i^k}{\partial n_i},$$

where the last sum is over all pairs of subdomains (i, j) that share an interface, and $a_i(u_i, v_i) = \int_{\Omega_i} (\nabla u \cdot \nabla v + \eta uv)\, dx$ is the energy bilinear form defined on subdomain Ω_i, so that $a_i(e_i^k, e_i^k) = \int_{\Omega_i} \eta |e_i^k|^2 + |\nabla e_i^k|^2\, dx \geq 0$ is the energy of the error on subdomain Ω_i. We now rewrite the product term as

$$e_i^k \frac{\partial e_i^k}{\partial n_i} = \frac{1}{4p_{ij}} \left[\left(\frac{\partial e_i^k}{\partial n_i} + p_{ij}e_i^k \right)^2 - \left(-\frac{\partial e_i^k}{\partial n_i} + p_{ij}e_i^k \right)^2 \right] =: \left(T_{+ij}^k\right)^2 - \left(T_{-ij}^k\right)^2,$$

where $T_{\pm ij}^k = \frac{1}{\sqrt{4p_{ij}}} (\pm \frac{\partial e_i^k}{\partial n_i} + p_{ij}e_i^k)$. Since $\frac{\partial e_j^k}{\partial n_i} = -\frac{\partial e_j^k}{\partial n_j}$ on Γ_{ij}, the interface condition in (3) can be written as $T_{+ij}^k = T_{-ji}^{k-1}$, which means

$$a_i(e_i^k, e_i^k) = \sum_{(i,j)\in E} \int_{\Gamma_{ij}} \left[\left(T_{+ij}^k\right)^2 - \left(T_{-ij}^k\right)^2 \right] ds = \sum_{(i,j)\in E} \int_{\Gamma_{ij}} \left[\left(T_{-ji}^{k-1}\right)^2 - \left(T_{-ij}^k\right)^2 \right] ds.$$

Thus,

$$a_i(e_i^k, e_i^k) + \sum_{(i,j)\in E} \int_{\Gamma_{ij}} \left(T_{-ij}^k\right)^2 ds = \sum_{(i,j)\in E} \int_{\Gamma_{ij}} \left(T_{-ji}^{k-1}\right)^2 ds. \tag{4}$$

If we sum (4) through all subdomains i, we get

$$\sum_{i=1}^N a_i(e_i^k, e_i^k) + \sum_{i=1}^N \sum_{(i,j)\in E} \int_{\Gamma_{ij}} \left(T_{-ij}^k\right)^2 ds = \sum_{i=1}^N \sum_{(i,j)\in E} \int_{\Gamma_{ij}} \left(T_{-ji}^{k-1}\right)^2 ds. \tag{5}$$

We can now sum (5) over k and simplify to get

$$\sum_{k=0}^K \sum_{i=1}^N a_i(e_i^k, e_i^k) + B^K = B^0, \tag{6}$$

where $B^k := \sum_{i=1}^{N} \sum_{(i,j)\in E} \int_{\Gamma_{ij}} \left(T^k_{-ij}\right)^2 ds \geq 0$. Since $B^K \geq 0$ and each $a_i(e_i^k, e_i^k) \geq 0$, we see that $\sum_{k=0}^{K} a_i(e_i^k, e_i^k) \leq B^0$ for all i and all K; hence $a_i(e_i^k, e_i^k) \to 0$ as $k \to \infty$ for all i. This implies that $\|e_i^k\|_{H^1(\Omega_i)} \to 0$ when $\eta > 0$, so $u_i \to u|_{\Omega_i}$ in the H^1 norm. A similar argument holds for $\eta = 0$. Note that the possible presence of cross points does not cause any difficulty in the proof, since they form a subset of measure zero in $\partial\Omega_i$ and thus do not contribute to the boundary terms when integrating by parts.

3 Finite Element Discretization

We now try to mimic Lions' proof in the finite element case. The finite element method uses the weak form of (2), i.e., we must multiply the PDE by a test function ϕ and integrate by parts. The problem becomes

$$\begin{aligned} &\textit{Find } u_i \in V^h \subset H^1(\Omega_i) \textit{ s.t. for all } \phi \in W^h \subset H_0^1(\Omega) \cap H^1(\Omega_i), \\ &\int_{\Omega_i} (\nabla\phi \cdot \nabla u_i^k + \eta\phi u_i^k) - \int_{\partial\Omega_i} \phi \frac{\partial u_i^k}{\partial n_i} = \int_{\Omega_i} \phi f. \end{aligned} \tag{7}$$

We now suppose that ϕ is a basis function corresponding to a degree of freedom along Γ_{ij}, whose support does not contain any cross points, see Fig. 1a To obtain an expression for $\int_{\partial\Omega_i} \phi \frac{\partial u_i^k}{\partial n_i}$, we multiply the interface condition by ϕ and integrate to get

$$\int_{\Gamma_{ij}} \phi\left(\frac{\partial u_i^k}{\partial n_i} + p u_i^k\right) = \int_{\Gamma_{ij}} \phi\left(\frac{\partial u_j^{k-1}}{\partial n_i} + p u_j^{k-1}\right). \tag{8}$$

Substituting into (7) gives

$$a_i(\phi, u_i^k) + \int_{\Gamma_{ij}} \phi p u_i^k - \int_{\Gamma_{ij}} \phi \frac{\partial u_j^{k-1}}{\partial n_i} = \int_{\Omega_i} \phi f + \int_{\Gamma_{ij}} \phi p u_j^{k-1}. \tag{9}$$

Thus, we are faced with the same problem of finding an expression for $\int_{\Gamma_{ij}} \phi \frac{\partial u_j^{k-1}}{\partial n_i}$. Fortunately, we can use the weak form of the PDE from Ω_j

$$a_j(\phi, u_j^{k-1}) - \int_{\partial\Omega_j} \phi \frac{\partial u_j^{k-1}}{\partial n_j} = \int_{\Omega_j} \phi f. \tag{10}$$

Since $n_i = -n_j$ on Γ_{ij}, adding (9) and (10) and rearranging gives

$$a_i(\phi, u_i^k) + \int_{\Gamma_{ij}} \phi p u_i^k = \int_{\Omega_i} \phi f - a_j(\phi, u_j^{k-1}) + \int_{\Gamma_{ij}} \phi p u_j^{k-1}, \tag{11}$$

which is just the usual block-Jacobi splitting of the stiffness matrix along Γ_{ij}.

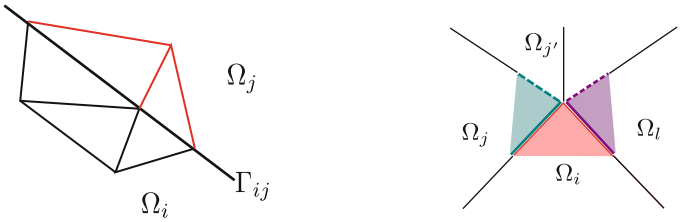


Fig. 1. Finite element discretization (**a**) without cross points and (**b**) with a cross point

Now assume that the support of ϕ contains cross points, see Fig. 1b. Here Ω_i is adjacent to two distinct subdomains Ω_j and Ω_l, $j \neq l$, and ϕ is non-zero on all three subdomains. Since the two parts of the interface, Γ_{ij} and Γ_{il}, must satisfy different interface conditions, we must separate $\int_{\partial\Omega_i} \phi \frac{\partial u_i^k}{\partial n}$ into contributions along Γ_{ij} and Γ_{il},

$$a_i(\phi, u_i^k) - \int_{\Gamma_{ij}} \phi \frac{\partial u_i^k}{\partial n_i} - \int_{\Gamma_{il}} \phi \frac{\partial u_i^k}{\partial n_i} = \int_{\Omega_i} \phi f.$$

The boundary term along Γ_{ij} can be replaced by the interface condition

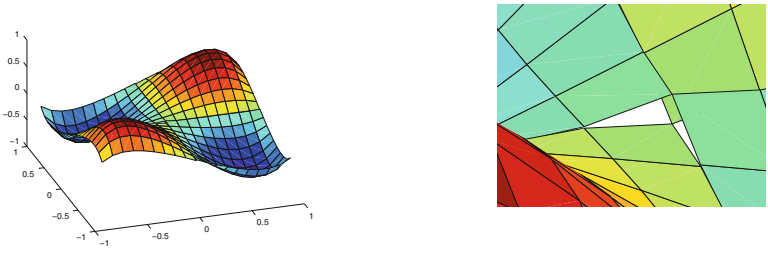
$$\int_{\Gamma_{ij}} \phi \left(\frac{\partial u_i^k}{\partial n_i} + p u_i^k\right) = \int_{\Gamma_{ij}} \phi \left(\frac{\partial u_j^{k-1}}{\partial n_i} + p u_j^{k-1}\right),$$

but now if we try to use the PDE on Ω_j to eliminate the term $\int_{\Gamma_{ij}} \phi \frac{\partial u_j^{k-1}}{\partial n_i}$, we would get

$$\int_{\Gamma_{ij}} \phi \frac{\partial u_j^{k-1}}{\partial n_j} = a_j(\phi, u_j^{k-1}) - \int_{\Gamma_{jj'}} \phi \frac{\partial u_j^{k-1}}{\partial n_j} - \int_{\Omega_j} \phi f,$$

so we get a new term representing the trace along $\Gamma_{jj'}$, where $\Omega_{j'}$ is another subdomain adjacent to j (see Fig. 1b). The same problem occurs when we try to eliminate the trace along Γ_{il}. Note that, in the discrete FEM setting, the Robin traces are integrated along a subset of $\partial\Omega_i$ of non-zero measure straddling both interfaces Γ_{ij} and Γ_{il}, and piecewise interface quantities are not available. Thus, the traces cannot be transmitted separately along Γ_{ij} and Γ_{il}, unlike in the continuous case; one must introduce extra unknowns to represent the piecewise Robin traces (integrated against a test function) for each subdomain at the cross point.

One way of circumventing the problem is to use mortar methods [1, 6], which are designed for non-conforming grids. In these methods, the interface conditions are imposed using mortar functions, which have one degree of freedom less at the ends of intervals. Thus, there is no equation at the cross point, and the problem of unavailable Robin traces goes away. However, since the interface conditions are only enforced weakly, the method does not generally converge to the exact solution of the global FEM problem, but rather to a discontinuous solution (Fig. 2) that is $O(h^p)$-accurate, where p is the order of the finite element method.



(a) (b)

Fig. 2. (**a**) The solution of $-\Delta u = f$ with four subdomains on $\Omega = [-1,1]^2$, with *right-hand side* $f(x,y) = \sin(xy)$. The interface conditions are imposed using a mortar space. (**b**) Discontinuity of the composite solution near the origin

4 Two Lagrange Multiplier and Primal-Dual Methods

If we want to formulate subdomain problems that are equivalent to the *discrete global* FEM problem, we need to introduce extra variables to represent the total Robin traces. Thus, at the cross point, we impose for each Ω_i

$$a_i(\phi, u_i^k) + \int_{\partial\Omega_i} p\phi \cdot u_i^k + \lambda_i^k = \int_{\Omega_i} \phi f, \tag{12}$$

where λ_i^k are Lagrange multipliers for ensuring consistency with the global problem. A cross point touching r subdomains requires r such Lagrange multipliers, so we also need r constraints to be satisfied at convergence:

- Continuity constraints ($r-1$ equations): at the cross point, we must have $u_1 = u_2 = \cdots = u_r$.
- PDE constraint (1 equation): if we sum (12) over the r subdomains and then subtract the global PDE $\sum_{i=1}^r a_i(\phi, u_i) = \int_\Omega \phi f$ from the result, we get

$$\sum_{i=1}^N \int_{\partial\Omega_i} p\phi u_i + \sum_{i=1}^N \lambda_i = 0.$$

This gives two types of algorithms:

1. Primal-Dual methods: the continuity constraints are enforced for every iteration. Thus, it suffices to introduce one extra variable (typically a coarse-grid basis function that has the value one at the cross point), and the PDE constraint is used as part of the coarse problem. This approach is similar to FETI-DP [3], except it is usually formulated with Neumann rather than Robin traces.
2. Two-Lagrange Multiplier methods: the λ_i^k are retained, but the u_i^k are eliminated using the PDE in the interior of the subdomains. This leads to a substructured problem formulated on the interface, which is then solved using a preconditioned Krylov method such as GMRES. This is known as the Two-Lagrange Multiplier (2LM) method and has been studied in detail in [8].

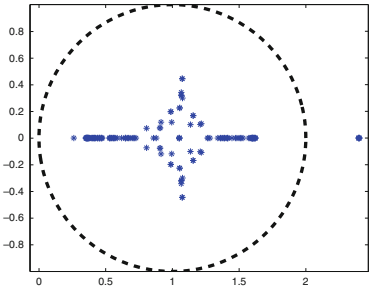


Fig. 3. Eigenvalues of the 2LM-preconditioned system for Poisson's equation ($\eta = 0$), using a 4×4 decomposition of the unit square with mesh size $h = 1/64$ and Robin parameter $p = C/\sqrt{h}$ for all interface nodes

Note that neither formulation is an exact discretization of (2) at cross points; thus, Lions' convergence analysis does not apply there. In fact, one can show [4] that the eigenvalues of the iteration matrix of the 2LM method may lie outside the unit disc when cross points are present, as seen in the 4×4 example shown in Fig. 3. In such cases, the method diverges. However, convergence can be restored if one uses Robin parameters with a different scaling at the cross points [4].

5 Conditions for Existence of Discrete Energy Estimates

To see what conditions are needed for Lions' estimates to hold in the discrete case, let us consider solving $-\Delta u = f$ on $\Omega = [-1,1]^2$ using P^1 finite elements on a structured triangular mesh. This yields the system $Au = f$, where A is identical to the matrix obtained from finite differences. If we now divide Ω into four subdomains corresponding to the four quadrants of the plane, then an optimized Schwarz method must solve

$$(A_i + L_i)u_i^k = g_i^k \qquad \text{on each } \Omega_i.$$

Here, A_i is the partially assembled stiffness matrix for Ω_i, L_i corresponds to transmission conditions, and g_i^k is a function of f and u_j^{k-1} for $j \neq i$. To define the *discrete* error function, let us write $u_i^* = u^*|_{\Omega_i}$, where u^* is the exact solution to $Au = f$. Then the error on Ω_i is $e_i^k = u_i^k - u_i^*$, with discrete energy $a_i(e_i^k, e_i^k) = (e_i^k)^T A_i e_i^k > 0$ whenever $e_i^k \neq 0$, since each subdomain touches a Dirichlet boundary. Now observe that

$$A_i e_i^k = A_i u_i^k - A_i u_i^* = A_i u_i^k - f_i \qquad \text{at interior nodes.}$$

Since the stencils of A_i and A coincide at interior nodes, we see that $A_i e_i^k$ must be zero away from the interfaces. Thus, we in fact have

$$a_i(e_i^k, e_i^k) = \sum_{v \in \partial\Omega_i \setminus \partial\Omega} e_i^k(v) \cdot (A_i e_i^k)(v) = \sum_{v \in \partial\Omega_i \setminus \partial\Omega} [(T_{+i}^k(v))^2 - (T_{-i}^k(v))^2],$$

where $T_{\pm i}^k(v)$ are the "Robin traces" at an interface point v:

$$T^k_{+i}(v) = \tfrac{1}{\sqrt{4p}}\big[(A_i e^k_i)(v) + p e^k_i(v)\big], \qquad T^k_{-i}(v) = \tfrac{1}{\sqrt{4p}}\big[-(A_i e^k_i)(v) + p e^k_i(v)\big].$$

Hence, if we let $T^k_{+i}(v) = T^{k-1}_{-j}(v)$ at every point v on the interface, then the energy estimate holds exactly the same way as in the continuous case, and we have convergence of the method. This allows us to deduce the correct interface conditions for v away from the cross point. Using the definition $e^k_i = u^k_i - u^*_i$, we have

$$(A_i(u^k_i - u^*_i))(v) + p(u^k_i(v) - u^*_i(v)) = -(A_j(u^{k-1}_j - u^*_j))(v) + p(u^{k-1}_j(v) - u^*_j(v)). \tag{13}$$

But since

$$(A_i u^*_i)(v) + (A_j u^*_j)(v) = f(v), \tag{14}$$

we can simplify (13) to get

$$(A_i u^k_i)(v) + p u^k_i(v) = f(v) - (A_j u^{k-1}_j)(v) + p u^{k-1}_j(v).$$

In other words, we need

$$(L_i u^k_i)(v) = p u^k_i(v), \qquad g^k_i(v) = f(v) - (A_j u^{k-1}_j)(v) + p u^{k-1}_j(v).$$

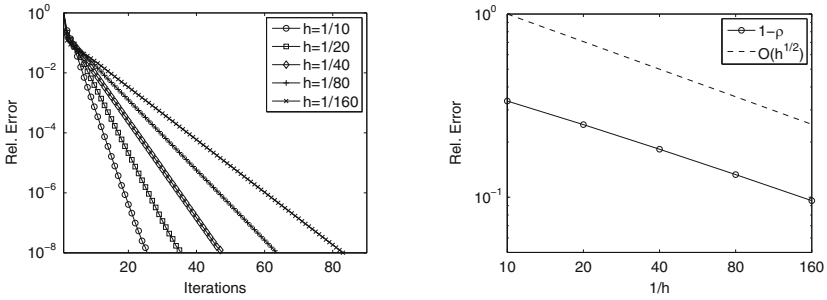
On the other hand, if v is a cross point, then (14) is no longer valid, since $f(v)$ is the sum of many subdomain contributions. Thus, it is in general impossible to find L_i and g^k_i such that the relation $T^k_{+i}(v) = T^{k-1}_{-j}(v)$ holds at the cross point for some j. In our model problem, however, the stencil at the cross point has a special form for the first and third quadrant:

$$(A_1 u^*_1)(0,0) = u^*(0,0) - \tfrac{1}{2}u^*(0,h) - \tfrac{1}{2}u^*(h,0),$$
$$(A_3 u^*_3)(0,0) = u^*(0,0) - \tfrac{1}{2}u^*(0,-h) - \tfrac{1}{2}u^*(-h,0).$$

Thus, we actually have $(A_1 u^*_1)(0,0) + (A_3 u^*_3)(0,0) = \frac{1}{2}f(0,0)$, a known quantity! A similar relation holds between Ω_2 and Ω_4, so it is actually possible to find transmission conditions at the cross point that satisfy the discrete energy estimate. For Ω_1, this reads

$$(A_1 u^k_1)(v) + p u^k_1(v) = \tfrac{1}{2}f(v) - (A_3 u^{k-1}_3)(v) + p u^{k-1}_3(v).$$

Figure 4 shows the convergence of the method for $p = \frac{\pi}{2\sqrt{h}}$, which gives the optimal contraction factor $\rho = 1 - O(\sqrt{h})$, just as in the two-subdomain case. Since the discrete energy estimate holds, the converged subdomain solutions always coincide with the *exact* discrete solution u^*, unlike in the mortar case. In general, discrete energy estimates can only be derived if *for every cross point v, its set of neighbors can be partitioned into disjoint pairs* (i, j) *such that* $(A_i u^*_i)(v) + (A_j u^*_j)(v) = f_{ij}(v)$ *can be calculated without knowing* u^*. For cross points with wide stencils or an odd number of neighbors, this is not possible. In such cases, the methods in Sect. 4 are still excellent choices in practice, but one cannot use Lions' estimates to deduce convergence for arbitrary positive Robin parameters p.



(a) (b)

Fig. 4. (**a**) Convergence for different grid spacing h; (**b**) Contraction rate versus h

Acknowledgments The authors would like to thank Laurence Halpern for all the stimulating discussions concerning energy estimates.

Bibliography

[1] C. Bernardi, Y. Maday, and A. T. Patera. A new non conforming approach to domain decomposition: The mortar element method. In H. Brezis and J.-L. Lions, editors, *Collège de France Seminar*. Pitman, 1994.

[2] B. Després. *Méthodes de décomposition de domaines pour les problèmes de propagation d'ondes en régime harmonique*. PhD thesis, Univ. Paris IX Dauphine, 1991.

[3] C. Farhat, M. Lesionne, P. Le Tallec, K. Pierson, and D. Rixen. FETI-DP: a dual-primal unified FETI method — part I: a faster alternative to the two-level FETI method. *Internat. J. Numer. Methods Engrg.*, 50:1523–1544, 2001.

[4] M. J. Gander and F. Kwok. Best Robin parameters for optimized Schwarz methods at cross points. *SIAM J. Sci. Comput.*, 34(4):A1849–A1879, 2012.

[5] M. J. Gander, L. Halpern, and F. Nataf. Optimized Schwarz waveform relaxation for the one dimensional wave equation. *SIAM J. Numer. Anal.*, 41:1643–1681, 2003.

[6] M. J. Gander, C. Japhet, Y. Maday, and F. Nataf. A new cement to glue nonconforming grids with Robin interface conditions: The finite element case. In *Domain Decomposition Methods in Science and Engineering, LNCSE 40*, pages 259–266. Springer Verlag, 2004.

[7] P.-L. Lions. On the Schwarz alternating method III: a variant for non-overlapping subdomains. In *Third international symposium on domain decomposition methods for partial differential equations*, pages 47–70, 1990.

[8] S. Loisel. Condition number estimates for the nonoverlapping optimized Schwarz method and the 2-Lagrange multiplier method for general domains and cross points. *Heriot-Watt Mathematics Report HWM11-5*, 1 Mar 2011.

[9] A. St-Cyr, D. Rosenberg, and S. D. Kim. Optimized Schwarz preconditioning for SEM based magnetohydrodynamics. In *Domain Decomposition Methods in Science and Engineering XVIII*, 2009.

Non Shape Regular Domain Decompositions: An Analysis Using a Stable Decomposition in H_0^1

Martin J. Gander[1], Laurence Halpern[2], and Kévin Santugini Repiquet[3]

[1] Université de Genève, Section de Mathématiques, Martin.Gander@unige.ch
[2] Université Paris 13, LAGA UMR 7539 CNRS halpern@math.univ-paris13.fr
[3] Université Bordeaux, IMB, UMR 5251 CNRS, INRIA, F-33400 Talence, France. Kevin.Santugini@math.u-bordeaux1.fr

Summary. In this paper, we establish the existence of a stable decomposition in the Sobolev space H_0^1 for domain decompositions which are not shape regular in the usual sense. In particular, we consider domain decompositions where the largest subdomain is significantly larger than the smallest subdomain. We provide an explicit upper bound for the stable decomposition that is independent of the ratio between the diameter of the largest and the smallest subdomain.

1 Introduction

One of the great success stories in domain decomposition methods is the invention and analysis of the additive Schwarz method by Dryja and Widlund in [2]. Even before the series of international conferences on domain decomposition methods started, Dryja and Widlund presented a variant of the historical alternating Schwarz method invented by Schwarz in [5] to prove the Dirichlet principle on general domains. This variant, called the additive Schwarz method, has the advantage of being symmetric for symmetric problems, and it also contains a coarse space component. In a fully discrete analysis in [2], Dryja and Widlund proved, based on a stable decomposition result for shape regular decompositions, that the condition number of the preconditioned operator with a decomposition into many subdomains only grows linearly as a function of $\frac{H}{\delta}$, where H is the subdomain diameter, and δ is the overlap between subdomains. This analysis inspired a generation of numerical analysts, who used these techniques in order to analyze many other domain decomposition methods, see the reference books [4, 6, 7], or the monographs [1, 8], and references therein.

The key assumption that the decomposition is shape regular is, however, often not satisfied in practice: because of load balancing, highly refined subdomains are often physically much smaller than subdomains containing less refined elements, and it is therefore of interest to consider domain decompositions that are only locally shape regular, i.e., domain decompositions where the largest subdomain can be considerably larger than the smallest subdomain, and therefore the subdomain

R. Bank et al. (eds.), *Domain Decomposition Methods in Science and Engineering XX*, Lecture Notes in Computational Science and Engineering 91,
DOI 10.1007/978-3-642-35275-1_57,

diameter and overlap parameters depend strongly on the subdomain index. In such a domain decomposition, the generic ratio $\frac{H}{\delta}$ from the classical convergence result of the additive Schwarz method can be given at least two different meanings: let H_i refer to the diameter of subdomain number i and δ_i refer to the width of the overlap around subdomain number i. Then in the classical convergence result from [2], one could replace the generic ratio $\frac{H}{\delta}$ by $\frac{\max_i(H_i)}{\min_i(\delta_i)}$, but this is likely to lead to a very pessimistic estimate for the condition number growth. The general analysis of the additive Schwarz method based on a shape regular decomposition does unfortunately not permit to answer the question if the condition number growth for a locally shape regular decomposition is in fact only linear in the quantity $\max_i(\frac{H_i}{\delta_i})$, which is much smaller than $\frac{\max_i(H_i)}{\min_i(\delta_i)}$ in the case of subdomains and overlaps of widely different sizes, a case of great interest in applications.

In [3], we established the existence of a stable decomposition in the continuous setting with an explicit upper bound and a quantitative definition of shape regularity in two spatial dimensions. The explicit upper bound is also linear in the generic quantity $\frac{H}{\delta}$, and the result is limited to shape regular domain decompositions where all subdomains have similar size and where the overlap width is uniform over all subdomains. Having explicit upper bounds, however, allows us now, using similar techniques, to establish the existence of a stable decomposition in the continuous setting with explicit upper bounds when $\max_i(H_i) \gg \min_i(H_i)$, and we provide an explicit upper bound which is linear in $\max_i(H_i/\delta_i)$ for problems in two spatial dimensions. To get this result, only a few of the inequalities established in [3] need to be reworked, and it would be very difficult to obtain such a result without the explicit upper bounds from the continuous analysis in [3].

We state first in Sect. 2 our main theorem along with the assumptions we make on the domain decomposition. We then prove the main theorem in Sect. 3 in two steps: first, we show in Lemma 1 how to construct the fine component in Sect. 3.1, which is an extension of the result [3, Theorem 4.6] for the case where subdomain sizes H_i and overlaps δ_i can strongly depend on the subdomain index i. The major contribution is however in the second step, presented in Lemma 2 in Sect. 3.2, where we show how to construct the coarse component in the case of strongly varying H_i and δ_i between subdomains. This result is a substantial generalization of [3, Lemma 5.7]. Using these two new results, and the remaining estimates from [3] which are still valid, we can prove our main theorem. We finally summarize our results in the conclusions in Sect. 4.

2 Geometric Parameters and Main Theorem

In the remainder of this paper, we always consider a domain decomposition that has the following properties:

- Ω is a bounded domain of $\mathbb{R}^2$.

- The $(U_i)_{1\le i\le N}$ are a non-overlapping domain decomposition of Ω, i.e., satisfy $\bigcup_{i=1}^N \overline{U}_i = \overline{\Omega}$ and $U_i \cap U_j = \emptyset$ when $i \neq j$. The U_i are bounded connected open sets of $\mathbb{R}^2$ and for all subdomains U_i the measure of $\overline{U}_i \setminus U_i$ is zero.
- We set $H_i := \mathrm{diam}(U_i)$.
- Two distinct subdomains U_i and U_j are said to be neighbors if $\overline{U}_i \cap \overline{U}_j \neq \emptyset$.
- For each subdomain U_i, let $\delta_i > 0$ be such that $2\delta_i \le \min_{j,\overline{U}_i\cap\overline{U}_j=\emptyset}(\mathrm{dist}(U_i,U_j))$. We set $\Omega_i := \{\boldsymbol{x} \in \Omega,\ \mathrm{dist}(\boldsymbol{x},U_i) < \delta_i\}$. The Ω_i form an overlapping domain decomposition of Ω. When subdomains U_i and U_j are neighbors, then the overlap between Ω_i and Ω_j is $\delta_i + \delta_j$ wide. The intersection $\Omega_i \cap \Omega_j$ is empty if and only if the distance between U_i and U_j is positive.
- We set $\delta_i^s = \min_{j\neq i,\overline{U}_i\cap\overline{U}_j\neq\emptyset} \delta_j$ and $\delta_i^l = \max_{j\neq i,\overline{U}_i\cap\overline{U}_j\neq\emptyset} \delta_j$.
- The domain decomposition has N_c colors: there exists a partition of $\mathbb{N}\cap[1,N]$ into N_c sets I_k such that $\Omega_i \cap \Omega_j$ is empty whenever $i \neq j$ and i and j belong to the same color I_k.
- $\mathscr{T}$ is a coarse triangular mesh of Ω: one node $\boldsymbol{x}_i$ per subdomain Ω_i (not counting the nodes located on $\partial\Omega$). By $P_1(\mathscr{T})$, we denote the standard finite element space of continuous functions that are piecewise linear over each triangular cell of $\mathscr{T}$.
- Let θ_{min} be the minimum of all angles of mesh $\mathscr{T}$.
- No node (including the nodes located on $\partial\Omega$) of the coarse mesh has more than K neighbors.
- Let d_i be the length of the largest edge originating from node $\boldsymbol{x}_i$ in the mesh $\mathscr{T}$.
- Let $H_{h,i}$ be the length of the shortest height through $\boldsymbol{x}_i$ of any triangle in the coarse mesh $\mathscr{T}$ that connects to $\boldsymbol{x}_i$. We also set $H'_{h,i}$ as the minimum of $H_{h,j}$ over i and its direct neighbors in mesh $\mathscr{T}$.
- We suppose that for each subdomain U_i, there exists $r_i > 0$ such that U_i is star-shaped with respect to any point in the ball $B(\boldsymbol{x}_i, r_i)$. We also suppose $r_i \le \frac{H_{h,i}}{4K+1}$ and $r_i \le H'_{h,i}/2$.
- We also assume the existence of both a pseudo normal $\boldsymbol{X}_i$ and of a pseudo curvature radius $\tilde{R}_i$ for the domain U_i, i.e., we suppose that for each U_i there exists an open layer L_i containing ∂U_i, a vector field $\boldsymbol{X}_i$ continuous on $L_i \cap \overline{U}_i$, $\mathscr{C}^\infty$ on $L_i \cap U_i$ such that $\mathrm{D}\boldsymbol{X}_i(\boldsymbol{x})(\boldsymbol{X}_i(\boldsymbol{x})) = 0$, $\|\boldsymbol{X}_i(\boldsymbol{x})\| = 1$, and $\varepsilon_0 > 0$ such that for all positive $\varepsilon < \varepsilon_0$ and for all $\hat{\boldsymbol{x}}$ in ∂U_i, $\hat{\boldsymbol{x}} + \varepsilon\boldsymbol{X}_i(\hat{\boldsymbol{x}}) \in U_i$ and $\hat{\boldsymbol{x}} - \varepsilon\boldsymbol{X}_i(\hat{\boldsymbol{x}}) \notin U_i$. We set, for all positive δ', $U_i^{\delta'} = \{\boldsymbol{x} \in U_i,\ \mathrm{dist}(\boldsymbol{x},\partial U_i) < \delta'\}$, and $V_i^{\delta'} = \{\hat{\boldsymbol{x}} + s\boldsymbol{X}_i(\hat{\boldsymbol{x}}), \hat{\boldsymbol{x}} \in \partial U_i, 0 < s < \delta'\}$. We assume there exist $\hat{R}_i > 0$, $\theta_{\boldsymbol{X}}$, $0 < \theta_{\boldsymbol{X}} \le \pi/2$, and δ_{0i}, $0 < \delta_{0i} \le \hat{R}_i \sin\theta_{\boldsymbol{X}}$ such that $V_i^{\hat{R}} \subset L_i \cap U_i$ and $U_i^{\delta'} \subset V^{\delta'/\sin\theta_{\boldsymbol{X}}}$ for all positive $\delta' \le \delta_{0i}$. Set $\tilde{R}_i := 1/\|\mathrm{div}\boldsymbol{X}_i\|_\infty$. We suppose $\delta_{0i} > \delta_i^l$.

We finally define, for all i, the linear form on $H_0^1(\Omega)$ by

$$\ell_i(u) := \frac{1}{\pi r_i^2}\int_{B(\boldsymbol{x}_i,r_i)} u(\boldsymbol{x})\mathrm{d}\boldsymbol{x} = \frac{1}{\pi}\int_{B(\boldsymbol{0},1)} u(\boldsymbol{x}_i + r_i\boldsymbol{y})\mathrm{d}\boldsymbol{y}.$$

We can now state our main theorem, namely the existence of a stable decomposition of $H_0^1(\Omega)$ whose upper bound is independent of $\frac{\max_i(H_i)}{\min_i(H_i)}$. This theorem therefore leads to a substantially sharper condition number estimate in the important case

of an only locally shape regular decomposition, and is a major improvement of [3, Theorem 5.12], which only considered shape regular decompositions, albeit at the continuous level, in contrast to [2].

Theorem 1. *For u in $H_0^1(\Omega)$, there exists a stable decomposition $(u_i)_{0\le i\le N}$ of u, i.e., $u=\sum_{i=0}^N u_i$, u_0 in $P_1(\mathscr{T})\cap H_0^1(\Omega)$ and $u_i\in H_0^1(\Omega_i)$ such that*

$$\sum_{i=0}^{N}\|\nabla u_i\|^2_{L^2(\Omega_i)}\le C\|\nabla u\|^2_{L^2(\Omega)},$$

where $C=2C_1+2(1+C_1)C_2$ and

$$C_1=\frac{1}{\tan\theta_{min}}\frac{\left(1+2\max_i(\frac{r_i}{H_{h,i}})\right)K(\frac{25}{6\pi}\max_i(\frac{d_i}{r_i})+2\pi)}{1-\left((2K+1)+(4K+1)\max_i(\frac{r_i}{H_{h,i}})\right)\max_i(\frac{r_i}{H_{h,i}})},$$

$$\begin{aligned}C_2=2+8\lambda_2^2(N_c-1)^2(1+\max_i\frac{\hat R_i}{\tilde R_i})\max_i\frac{\delta_i^l}{\delta_i^s}\max_i\frac{\hat R_i}{\delta_i^s\sin\theta_{\boldsymbol{X}}}\\ +\frac{8}{3}\lambda_2^2(N_c-1)^2(1+\max_i\frac{\hat R_i}{\tilde R_i})\max_i\frac{\delta_i^l}{\delta_i^s}\max_i\frac{r_i^2}{\delta_i^s\hat R_i\sin\theta_{\boldsymbol{X}}}\times\\ \times\max_i\left(\left(\left(\frac{H_i^2}{r_i^2}+\frac{1}{2}\right)^{\frac14}+\frac{H_i}{\sqrt[4]{2}r_i}\right)^4-\frac12-\frac{H_i^2}{r_i^2}-\frac{H_i^4}{2r_i^4}\right),\end{aligned}$$

with λ_2 a universal constant depending only on the dimension, and being smaller than 6 in the two dimensional case we consider here.

Note that the condition $r_i\le\frac{H_{h,i}}{4K+1}$ implies that the denominator of C_1 is positive. The value of C_2 is also always positive.

3 Proof of Theorem 1

The proof is based on the continuous analysis in [3], but two results must be adapted to the situation of only locally shape regular decompositions: we first show in Sect. 3.1 how to construct the fine component, which is a technical extension of the result [3, Theorem 4.6] for the case where subdomain sizes H_i and overlaps δ_i can strongly depend on the subdomain index i. Second, we explain in Sect. 3.2 the construction of the coarse component in the case of strongly varying H_i and δ_i between subdomains, which is a non-trivial generalization of [3, Lemma 5.7]. With these two new results, and the remaining estimates from [3], the proof can be completed.

3.1 Constructing the Fine Component

We begin by establishing a stable decomposition when there is no coarse mesh.

Lemma 1. *Let u be in $H^1_0(\Omega)$. Then, there exist $(u_i)_{1\le i\le N}$, u_i in $H^1_0(\Omega_i)$ such that $u=\sum_{i=1}^N u_i$, and*

$$\sum_{i=1}^N \|\nabla u_i\|^2_{L^2(\Omega)} \le 2\|\nabla u\|^2_{L^2(\Omega)} + 8\lambda_2^2(N_c-1)^2\left(\sum_{i=1}^N (1+\frac{\hat R_i}{\tilde R_i})\frac{\delta_i^l}{\delta_i^s}\frac{\hat R_i}{\delta_i^s\sin\theta_{\boldsymbol{X}}}\|\nabla u\|^2_{L^2(U_i)}\right) + 8\lambda_2^2(N_c-1)^2\left(\sum_{i=1}^N (1+\frac{\hat R_i}{\tilde R_i})\frac{\delta_i^l}{\delta_i^s}\frac{1}{\delta_i^s\hat R_i\sin\theta_{\boldsymbol{X}}}\|u\|^2_{L^2(U_i)}\right), \tag{1}$$

where λ_2 is the universal constant of Theorem 1. We further have, for all $\eta>0$,

$$\begin{aligned}\sum_{i=1}^N \|\nabla u_i\|^2_{L^2(\Omega)} \le{}& 2\|\nabla u\|^2_{L^2(\Omega)} + 8\lambda_2^2(N_c-1)^2\sum_{i=1}^N (1+\frac{\hat R_i}{\tilde R_i})\frac{\delta_i^l}{\delta_i^s}\frac{\hat R_i}{\delta_i^s\sin\theta_{\boldsymbol{X}}}\|\nabla u\|^2_{L^2(U_i)}\\ &+\frac{8(1+\eta)}{3}\lambda_2^2(N_c-1)^2\sum_{i=1}^N (1+\frac{\hat R_i}{\tilde R_i})\frac{\delta_i^l}{\delta_i^s}\frac{r_i^2}{\delta_i^s\hat R_i\sin\theta_{\boldsymbol{X}}}\times\\ &\times\left(\left(\left(\frac{H_i^2}{r_i^2}+\frac12\right)^{\frac14}+\frac{H_i}{\sqrt[4]{2}r_i}\right)^4-\frac12-\frac{H_i^2}{r_i^2}-\frac{H_i^4}{2r_i^4}\right)\|\nabla u\|^2_{L^2(U_i)}\\ &+8(1+\frac1\eta)\pi\lambda_2^2(N_c-1)^2\sum_{i=1}^N (1+\frac{\hat R_i}{\tilde R_i})\frac{\delta_i^l}{\delta_i^s}\frac{H_i^2}{\delta_i^s\hat R_i\sin\theta_{\boldsymbol{X}}}|\ell_i(u)|^2.\end{aligned} \tag{2}$$

Proof. We follow the proof of [3, Theorem 4.6]. Let ρ be a $\mathscr{C}^\infty$ non-negative function whose support is included in the closed unit ball of $\mathbb{R}^2$ and whose L^1 norm is 1. Let $\rho_\varepsilon(\boldsymbol{x})=\rho(\boldsymbol{x}/\varepsilon)/\varepsilon^2$ for all $\varepsilon>0$. Let h_i be the characteristic function of the set $\{\boldsymbol{x}\in\mathbb{R}^2, \mathrm{dist}(\boldsymbol{x},U_i)<\delta_i/2\}$. Let $\phi_i=\rho_{\delta_i/2}*h_i$. The function ϕ_i is equal to 1 inside U_i, vanishes outside of $\{\boldsymbol{x}\in\mathbb{R}^2, \mathrm{dist}(\boldsymbol{x},U_i)<\delta_i\}$, and $\|\nabla\phi_i\|_{L^\infty(\mathbb{R}^2)}\le 2\|\nabla\rho\|_{L^1(\mathbb{R}^2;(\mathbb{R}^2,\|\cdot\|_2))}/\delta_i$. Here, $\|\nabla\rho\|_{L^1(\mathbb{R}^2;(\mathbb{R}^2,\|\cdot\|_2))}$ means $\int_{\mathbb{R}^2}\sqrt{\sum_{i=1}^2|\partial_i\rho|^2}\mathrm{d}\boldsymbol{x}$.

For i in $\mathbb{N}\cap[1,N]$, let $\psi_i=\phi_i\prod_{k=1}^{i-1}(1-\phi_k)$. We have $0\le\psi_i\le1$, ψ_i zero in $\Omega\setminus\Omega_i$ and $\sum_i\psi_i=1$ in Ω. Set $u_i=\psi_i u$. The function u_i is in $H^1_0(\Omega_i)$ and $u=\sum_i u_i$. Following the proof of [3, Lemma 4.3], we get $\sum_{i=1}^N\|\nabla\psi_i(\boldsymbol{x})\|_2^2\le 2(N_C-1)\sum_{i=1}^N\|\nabla\phi_i(\boldsymbol{x})\|_2^2$. Therefore, for all $\boldsymbol{x}$ in Ω,

$$\sum_{i=1}^N\|\nabla\psi_i(\boldsymbol{x})\|_2^2\le 8(N_c-1)\|\nabla\rho\|^2_{L^1(\mathbb{R}^2;(\mathbb{R}^2,\|\cdot\|_2))}\sum_{i=1}^N\frac{\mathbb{1}_{\Omega_i\setminus U_i}(\boldsymbol{x})}{\delta_i^2},$$

where $\mathbb{1}_{\mathscr{O}}$ is the indicator function for the set $\mathscr{O}$. Since $\sum_i\|\nabla u_i\|^2_{L^2(\Omega)}\le 2\|\nabla u\|^2_{L^2(\Omega)}+2\int_\Omega|u(\boldsymbol{x})|^2\sum_i|\nabla\psi_i(\boldsymbol{x})|^2\mathrm{d}\boldsymbol{x}$, we get

$$\sum_{i=1}^N\|\nabla u_i\|^2_{L^2(\Omega)}\le 2\|\nabla u\|^2_{L^2(\Omega)}+4\lambda_2^2(N_c-1)^2\sum_{i=1}^N\int_{U_i}\mathbb{1}_{\{\mathrm{dist}(\boldsymbol{x},\partial U_i)<\delta_i^l\}}\frac{|u(\boldsymbol{x})|^2}{(\delta_i^s)^2}\mathrm{d}\boldsymbol{x},$$

with $\lambda_2 := 2\|\nabla\rho\|_{L^1(\mathbb{R}^2;(\mathbb{R}^2,\|\cdot\|_2))}$. Using the $W^{1,1}(\mathbb{R}^2)$ function $\rho(\boldsymbol{x}) = 1 - \|\boldsymbol{x}\|_2$, we obtain the estimate $\lambda_2 = 6$. To get (1), we apply Lemma 4.5 in [3] to each U_i, and to obtain (2), we apply Lemma 5.10 from the same reference. □

To obtain a stable decomposition with a coarse component, we want to construct u_0 in $P_1(\mathscr{T})$ such that for all i, $\ell_i(u_0) = \ell_i(u)$.

3.2 Constructing the Coarse Component

To construct u_0, we follow the ideas of [3, Sect. 5.2]. First, we define a special norm.

Definition 1. *Let $\mathscr{T}$ be the coarse mesh of the domain Ω. Let $\mathscr{B}'$ be the set of indices of the nodes of $\mathscr{T}$ located on the boundary*[4] *$\partial\Omega$. Let $\mathscr{B}$ be the set of the indices of the nodes that are neighbors to the nodes with index in $\mathscr{B}'$. Let $\mathscr{V}$ be the set of pairs of indices of neighboring nodes in $\mathscr{T}$ which are not on $\partial\Omega$. We define*

$$\|\cdot\|_{\mathscr{V},\mathscr{B}} : \mathbb{R}^N \to \mathbb{R}^+,$$
$$\boldsymbol{y} \mapsto \sqrt{\sum_{(i,j)\in\mathscr{V}} |y_i - y_j|^2 + \sum_{i\in\mathscr{B}} |y_i|^2}.$$

When u is in $P_1(\mathscr{T}) \cap H_0^1(\Omega)$, set $\|u\|_{\mathscr{V},\mathscr{B}} := \|(u(\boldsymbol{x}_i))_{1\le i\le N}\|_{\mathscr{V},\mathscr{B}}$, where the $\boldsymbol{x}_i$ are the interior nodes of the mesh $\mathscr{T}$.

Lemma 2. *For u in $H_0^1(\Omega)$, there exists u_0 in $P_1(\mathscr{T}) \cap H_0^1(\Omega)$ such that, for all i in $\{1,\dots,N\}$, $\ell_i(u_0) = \ell_i(u)$ and*

$$\|\nabla u_0\|^2_{L^2(\Omega)} \le \frac{1}{\tan\theta_{min}} \frac{\left(1 + 2\max_i(\frac{r_i}{H_{h,i}})\right) K\left(\frac{25}{6\pi}\max_i(\frac{d_i}{r_i}) + 2\pi\right)}{1 - \left((2K+1) + (4K+1)\max_i(\frac{r_i}{H_{h,i}})\right)\max_i(\frac{r_i}{H_{h,i}})} \|\nabla u\|^2_{L^2(\Omega)}.$$

Proof. The results of [3, Lemmas 5.6 and 5.8] stand without modifications. Therefore u_0 exists, and we have

$$\|\nabla u_0\|^2_{L^2(\Omega)} \le \frac{1}{\tan\theta_{min}} \frac{1 + 2\max_i(\frac{r_i}{H_{h,i}})}{1 - \left((2K+1) + (4K+1)\max_i(\frac{r_i}{H_{h,i}})\right)\max_i(\frac{r_i}{H_{h,i}})} \|(\ell_i(u))_{1\le i\le N}\|^2_{\mathscr{V},\mathscr{B}}.$$

Note that the condition $r_i \le \frac{H_{h,i}}{4K+1}$ implies the second denominator in the above equation is positive.

It remains to compare $\|u\|^2_{\mathscr{V},\mathscr{B}}$ and $\|\nabla u\|^2_{L^2(\Omega)}$. We need to adapt the proof of [3, Lemma 5.7]. We can suppose without any loss of generality that u is in $\mathscr{C}^\infty(\overline{\Omega})$. Let i, j in $\{1,\dots,N\}$ be indices of neighboring nodes of $\mathscr{T}$. Let $\boldsymbol{d}_{ij} = \boldsymbol{x}_i - \boldsymbol{x}_j$, and $d_{ij} = \|\boldsymbol{d}_{ij}\|$. We have for all $(i,j) \in \mathscr{V}$

[4] Because of the homogenous Dirichlet condition on the boundary $\partial\Omega$, the nodes whose indices are in $\mathscr{B}'$ are not associated to a degree of freedom, therefore $\mathscr{B}'$ and $\{1,\dots,N\}$ have empty intersection.

$$
\begin{aligned}
|\ell_i(u)-\ell_j(u)|^2 &= \frac{1}{\pi^2}\left(\int_{B(\mathbf{0},1)}(u(\boldsymbol{x}_i+r_i\boldsymbol{y})-u(\boldsymbol{x}_j+r_j\boldsymbol{y}))\mathrm{d}\boldsymbol{y}\right)^2\\
&\le \frac{1}{\pi}\int_{B(\mathbf{0},1)}\int_0^1 \|\nabla u\big(t(\boldsymbol{x}_i+r_i\boldsymbol{y})+(1-t)(\boldsymbol{x}_j+r_j\boldsymbol{y})\big)\|_2^2\|\boldsymbol{x}_i-\boldsymbol{x}_j+(r_i-r_j)\boldsymbol{y}\|_2^2\mathrm{d}t\mathrm{d}\boldsymbol{y}\\
&\le \frac{(d_{ij}+|r_i-r_j|)^2}{\pi}\int_{B(\mathbf{0},1)}\int_0^1 \|\nabla u\big(t(\boldsymbol{x}_i+r_i\boldsymbol{y})+(1-t)(\boldsymbol{x}_j+r_j\boldsymbol{y})\big)\|_2^2\mathrm{d}t\mathrm{d}\boldsymbol{y}\\
&\le \frac{(d_{ij}+|r_i-r_j|)^2}{\pi}\int_{T_{i,j}}\|\nabla u(\boldsymbol{y}')\|_2^2\int_0^1\frac{\mathbb{1}_{\{\|\boldsymbol{y}'-t\boldsymbol{x}_i-(1-t)\boldsymbol{x}_j\|\le tr_i+(1-t)r_j\}}}{(tr_i+(1-t)r_j)^2}\mathrm{d}t\mathrm{d}\boldsymbol{y}',
\end{aligned}
$$

where the tube $T_{i,j}$ is the convex hull of $B(\boldsymbol{x}_i,r_i)\cup B(\boldsymbol{x}_j,r_j)$. We get

$$
\begin{aligned}
&\max_{\boldsymbol{y}'\in\mathbb{R}^2}\int_0^1\frac{\mathbb{1}_{\{\|\boldsymbol{y}'-t\boldsymbol{x}_i-(1-t)\boldsymbol{x}_j\|\le tr_i+(1-t)r_j\}}}{(tr_i+(1-t)r_j)^2}\mathrm{d}t\\
&\quad=\max_{(s,s')\in\mathbb{R}^2}\int_0^1\frac{\mathbb{1}_{\{\sqrt{(s-td_{ij})^2+s'^2}\le tr_i+(1-t)r_j\}}}{(tr_i+(1-t)r_j)^2}\mathrm{d}t\\
&\quad=\max_{s\in[-r_j,d_{ij}+r_i]}\int_0^1\frac{\mathbb{1}_{\{|s-td_{ij}|\le tr_i+(1-t)r_j\}}}{(tr_i+(1-t)r_j)^2}\mathrm{d}t\\
&\quad\le\max_{s\in[-r_j,d_{ij}+r_i]}\int_{\frac{s-r_j}{d_{ij}+(r_i-r_j)}}^{\frac{s+r_j}{d_{ij}-(r_i-r_j)}}\frac{1}{(tr_i+(1-t)r_j)^2}\mathrm{d}t\\
&\quad=\max_{s\in[-r_j,d_{ij}+r_i]}-\frac{1}{r_i-r_j}\left[\frac{1}{(tr_i+(1-t)r_j)}\right]_{\frac{s-r_j}{d_{ij}+(r_i-r_j)}}^{\frac{s+r_j}{d_{ij}-(r_i-r_j)}}\\
&\quad=\max_{s\in[-r_j,d_{ij}+r_i]}\left(\frac{2}{d_{ij}r_j+s(r_i-r_j)}\right)\\
&\quad=\frac{2}{\min(r_i,r_j)(d_{ij}-|r_i-r_j|)}.
\end{aligned}
$$

Since $d_{ij}\ge H_{h,i}\ge 4\max(r_i,r_j)$, we have

$$
|\ell_i(u)-\ell_j(u)|^2\le\frac{25d_{ij}}{6\pi\min(r_i,r_j)}\|\nabla u\|^2_{L^2(T_{ij})}. \tag{3}
$$

If i is in the boundary set of the coarse mesh, then the node $\boldsymbol{x}_i$ is neighbor to a node $\boldsymbol{x}_{i'}$ located on $\partial\Omega$. Note that i' lies outside of the range $\{1,\dots,N\}$. Using [3, Eqs. (5.7) and (5.9)], we get

$$
\sum_{i\in\mathscr{B}}|\ell_i(u)|^2\le\left(\sum_{i\in\mathscr{B}}\frac{4\|\boldsymbol{x}_i-\boldsymbol{x}_{i'}\|}{\pi r_i}\int_{T_i'}\|\nabla u(\boldsymbol{x})\|^2\mathrm{d}\boldsymbol{x}\right)+2K\pi\|\nabla u\|^2_{L^2(\Omega)}, \tag{4}
$$

where T_i' is the convex hull of $B(\boldsymbol{x}_i,r_i)\cup B(\boldsymbol{x}_{i'},r_i)$. We sum inequality (3) over all i,j in the neighbor set and combine the resulting inequality with Eq. (4). Since

$\max(r_i, r_j) \leq H'_{h,i}/2 \leq \min(H_{h,i}, H_{h,j})/2$, no point can belong to more than K tubes $T_{i,j}$ or T'_i. Therefore, $\|(\ell_i(u))_{1\leq i\leq N}\|^2_{\mathscr{V},\mathscr{B}} \leq K\big(25\max_i(d_i/r_i)/(6\pi)+2\pi\big)\|\nabla u\|^2_{L^2(\Omega)}$. This concludes the proof. □

To prove Theorem 1, we use Lemma 2 to construct the coarse component u_0. We then apply Lemma 1 to $u - u_0$ to get the fine components u_i. The terms in $\ell_i(u)$ vanish.

4 Conclusion

We have proved the existence of a stable decomposition of the Sobolev space $H^1_0(\Omega)$ in the presence of a coarse mesh when the domain decomposition is only guaranteed to be locally shape regular. We provided an explicit upper bound for the stable decomposition that depends neither on $\max_i(H_i)/\min_i(H_i)$, nor on the number of subdomains. This would not have been possible without the explicit upper bounds provided in [3]. This shows that deriving such explicit upper bounds can be important for problems arising naturally in applications, e.g., load balanced domain decompositions with local refinement.

Bibliography

[1] Tony F. Chan and Tarek P. Mathew. Domain decomposition algorithms. In *Acta Numerica 1994*, pages 61–143. Cambridge University Press, 1994.

[2] Maksymilian Dryja and Olof B. Widlund. An additive variant of the Schwarz alternating method for the case of many subregions. Technical Report 339, also Ultracomputer Note 131, Department of Computer Science, Courant Institute, 1987.

[3] Martin J. Gander, Laurence Halpern, and Kévin Santugini-Repiquet. Continuous Analysis of the Additive Schwarz Method: a Stable Decomposition in H^1. *Submitted*, 2011. URL http://hal.archives-ouvertes.fr/hal-00462006/fr/.

[4] Alfio Quarteroni and Alberto Valli. *Domain Decomposition Methods for Partial Differential Equations*. Oxford Science Publications, 1999.

[5] Hermann A. Schwarz. Über einen Grenzübergang durch alternierendes Verfahren. *Vierteljahrsschrift der Naturforschenden Gesellschaft in Zürich*, 15: 272–286, May 1870.

[6] Barry F. Smith, Petter E. Bjørstad, and William Gropp. *Domain Decomposition: Parallel Multilevel Methods for Elliptic Partial Differential Equations*. Cambridge University Press, 1996.

[7] Andrea Toselli and Olof Widlund. *Domain Decomposition Methods - Algorithms and Theory*, volume 34 of *Springer Series in Computational Mathematics*. Springer, 2004.

[8] Jinchao Xu. Iterative methods by space decomposition and subspace correction. *SIAM Review*, 34(4):581–613, December 1992.

Overlapping Domain Decomposition: Convergence Proofs

Minh-Binh Tran

Basque Center for Applied Mathematics, Alameda de Mazarredo, 14 E-48009 Bilbao, Basque Country - Spain. binh@bcamath.org

1 Introduction

During the last two decades many domain decomposition algorithms have been constructed and lot of techniques have been developed to prove the convergence of the algorithms at the continuous level. Among the techniques used to prove the convergence of classical Schwarz algorithms, the first technique is the maximum principle used by Schwarz. Adopting this technique M. Gander and H. Zhao proved a convergence result for n-dimensional linear heat equation in [4]. The second technique is that of the orthogonal projections, used by P. L. Lions in [7], and his convergence results are for linear Laplace equation and linear Stokes equation. In the same paper, P. L. Lions also proved that the Schwarz sequences for linear elliptic equations are related to classical minimization methods over product spaces and this technique was then used by L. Badea in [1] for nonlinear monotone elliptic problems. Another technique is the Fourier and Laplace transforms used in the papers [3, 5] for some 1-dimensional evolution equations, with constant coefficients. In [10, 11], S. H. Lui used the idea of upper-lower solutions methods to study the convergence problem for some PDEs, with initial guess to be an upper or lower solution of the equations and monotone iterations. For nonoverlapping optimized Schwarz methods, P. L. Lions in [8] proposed to use an energy estimate argument to study the convergence of the algorithm. The energy estimate technique was then developed in [2] for Helmholtz equation and it has then become a very powerful tool to study nonoverlapping problems. J.-H. Kimn in [6] proved the convergence of an overlapping optimized Schwarz method for Poisson's equation with Robin boundary data and S. Loisel and D. B. Szyld in [9] extended the technique of J.-H. Kimn to linear symmetric elliptic equation. Another technique is to use semiclassical analysis, which works for overlapping optimized Schwarz methods with rectangle subdomains, linear advection diffusion equations on the half plane (see [12]). This paper is devoted to the study of the convergence of Schwarz methods at the continuous level. We give a sketch of the proof of the convergence of optimized Schwarz methods for semilinear parabolic equations, with multiple subdomains. Complete convergence proofs for both classical

R. Bank et al. (eds.), *Domain Decomposition Methods in Science and Engineering XX*, Lecture Notes in Computational Science and Engineering 91,
DOI 10.1007/978-3-642-35275-1_58,

and optimized Schwarz methods, both semilinear parabolic and elliptic equations, with multiple subdomains could be found in [13].

2 Convergence for Semilinear Parabolic Equations

Consider the following parabolic equation

$$\begin{cases} \frac{\partial u}{\partial t}(x,t) - \sum_{i,j=1}^{n} a_{i,j}(x)\frac{\partial^2 u}{\partial x_i \partial x_j}(x,t) + \sum_{i=1}^{n} b_i(x)\frac{\partial u}{\partial x_i}(x,t) \\ \qquad\qquad +c(x)u(x,t) = F(x,t,u(x,t)), \text{ in } \Omega \times (0,\infty), \\ u(x,t) = g(x,t), \text{ on } \partial\Omega \times (0,\infty), \\ u(x,0) = g(x,0), \text{ on } \Omega, \end{cases} \tag{1}$$

where Ω is a bounded and smooth enough domain in $\mathbb{R}^n$. The following conditions are imposed on 1).
(A1) For all i, j in $\{1,\ldots,I\}$, $a_{i,j}(x) = a_{j,i}(x)$. There exist strictly positive numbers λ, Λ such that $A = (a_{i,j}(x)) \geq \lambda I$ in the sense of symmetric positive definite matrices and $a_{i,j}(x) < \Lambda$ in Ω.
(A2) The functions $a_{i,j}$, b_i, c are in $C^\infty(\mathbb{R}^n)$ and g is in $C^\infty(\mathbb{R}^{n+1})$.
(A3) There exists $C > 0$, such that $\forall\ t \in \mathbb{R}$, $\forall\ x \in \mathbb{R}^n$, $|F(x,t,z) - F(x,t,z')| \leq C|z - z'|$, $\forall\ z, z' \in \mathbb{R}$. We now describe the way that we decompose the domain Ω: The domain Ω is divided into I smooth overlapping subdomains $\{\Omega_l\}_{l\in\{1,I\}}$:

$$(\partial\Omega_l \backslash \partial\Omega) \cap (\partial\Omega_{l'} \backslash \partial\Omega) = , \ \forall\ l, l' \in \{1,\ldots,I\}, \ l \neq l';$$

$$\forall l \in \{1,\ldots,I\}, \forall l', l'' \in J_l, l'' \neq l', \quad \Omega_{l'} \cap \Omega_{l''} = ,$$

where

$$J_l = \{l' | \Omega_{l'} \cap \Omega_l \neq \};$$

$$\cup_{l=1}^{n} \Omega_l = \Omega.$$

This decomposition means that we do not consider cross-points in this paper.
Denote by $\Gamma_{l,l'}$, for $l' \in J_l$, the set $(\partial\Omega_l \backslash \partial\Omega) \cap \overline{\Omega}_{l'}$. The transmission operator $\mathfrak{B}_{l,l'}$ is of Robin type $\mathfrak{B}_{l,l'} v = \sum_{i,j=1}^{n} a_{i,j}\frac{\partial v}{\partial x_i} n_{l,l',j} + p_{l,l'} v$ and $n_{l,l',j}$ is the j-th component of the outward unit normal vector of $\Gamma_{l,l'}$; $p_{l,l'}$ is positive and belongs to $L^\infty(\Gamma_{l,l'})$. The iterate $\#k$ in the l-th domain, denoted by u_l^k of the Schwarz waveform relaxation algorithm is defined by:

$$\begin{cases} \frac{\partial u_l^k}{\partial t} - \sum_{i,j=1}^{n} a_{i,j}\frac{\partial^2 u_l^k}{\partial x_i \partial x_j} + \sum_{i=1}^{n} b_i \frac{\partial u_l^k}{\partial x_i} + c u_l^k = F(t,x,u_l^k), \text{ in } \Omega_l \times (0,\infty), \\ \mathfrak{B}_{l,l'} u_l^k = \mathfrak{B}_{l,l'} u_{l'}^{k-1}, \text{ on } \Gamma_{l,l'} \times (0,\infty), \forall l' \in J_l, \end{cases} \tag{2}$$

where

$$u_l^k(x,t) = g(x,t) \text{ on } (\partial\Omega_l \cap \partial\Omega) \times (0,\infty), \quad u_l^k(x,0) = g(x,0) \text{ in } \Omega_l.$$

The initial guess u^0 is bounded in $C^\infty(\overline{\Omega \times (0,\infty)})$; and at step 0, the Eq. (2) is solved with boundary data

$$\mathfrak{B}_{l,l'} u_l^1(x,t) = u^0(x,t) \text{ on } \Gamma_{l,l'} \times (0,\infty), \forall l' \in J_l.$$

A compatibility condition on $u^0(x,t)$ is also assumed

$$\mathfrak{B}_{l,l'} g(x,0) = u^0(x,0) \text{ on } \Gamma_{l,l'}, \forall l' \in J_l.$$

By an induction argument, the algorithm is well-posed. Let e_l^k be $u_l^k - u$

$$\begin{cases} \frac{\partial e_l^k}{\partial t} - \sum_{i,j=1}^n a_{i,j}(x) \frac{\partial^2 e_l^k}{\partial x_i \partial x_j} + \sum_{i=1}^n b_i(x) \frac{\partial e_l^k}{\partial x_i} \\ \qquad + c(x) e_l^k = F(t,x,u_l^k) - F(t,x,u), & \text{in } \Omega_l \times (0,\infty), \\ \mathfrak{B}_{l,l'} e_l^k(x,t) = \mathfrak{B}_{l,l'} e_{l'}^{k-1}(x,t), & \text{on } \Gamma_{l,l'} \times (0,\infty), \forall l' \in J_l. \end{cases} \tag{3}$$

Moreover,

$$e_l^k(x,t) = 0 \text{ on } (\partial\Omega_l \cap \partial\Omega) \times (0,\infty), \quad e_l^k(x,0) = 0 \text{ in } \Omega_l.$$

For any function f in $L^2(0,\infty)$, define

$$\int_0^\infty f(x) \exp(-yx) dx.$$

For any fixed positive number α, define

$$|f|_\alpha = \sup_{\alpha' > \alpha} \left[\int_{\alpha'}^{\alpha'+1} \left(\int_0^\infty f(x) \exp(-yx) dx \right)^2 dy \right]^{\frac{1}{2}},$$

and

$$\mathbb{L}^2_\alpha(0,\infty) = \{ f \ : \ f \in L^2(0,\infty), |f|_\alpha < \infty \}.$$

Thus $(\mathbb{L}^2_\alpha(0,\infty), |.|_\alpha)$ is a normed subspace of $L^2(0,\infty)$.

Theorem 1. *Consider the Schwarz algorithm with Robin transmission conditions and the initial guess u^0 in $C_c^\infty(\overline{\Omega \times (0,\infty)})$. There exists a constant α large enough such that*

$$\lim_{k \to \infty} \sum_{l=1}^I \int_{\Omega_l} |e_l^k|_\alpha^2 dx = 0.$$

Proof. Let g_l be a function bounded and greater than 1 in $C^\infty(\mathbb{R}^n, \mathbb{R})$, α be a positive constant, we define

$$\Phi_l^k(x) := \left(\int_0^\infty e_l^k \exp(-\alpha t) dt \right) g_l(x),$$

then $\Phi_l^k(x)$ belongs to $H^1(\Omega_l)$. Let B_i^l and C^l be functions in $L^\infty(\mathbb{R}^n)$ defined by

$$B_i^l := b_i + \sum_{j=1}^n \left(a_{i,j} \frac{\partial_j g_l}{g_l} \right),$$

$$C^l = \left[\frac{\alpha}{2} + \sum_{i,j=1}^n \left(-a_{i,j} \frac{2\partial_i g_l \partial_j g_l}{(g_l)^2} - \partial_j a_{i,j} \frac{\partial_i g}{g} + a_{i,j} \frac{\partial_{i,j} g_l}{g_l} \right) - \sum_{i=1}^n b_i \frac{\partial_i g_l}{g_l} \right].$$

Define

$$\mathfrak{L}_{lR}(\Phi_l^k) = -\sum_{i,j=1}^n \partial_j (a_{i,j} \partial_i \Phi_l^k) + \sum_{i=1}^n B_i^l \partial_i \Phi_l^k + C^l \Phi_l^k + \left\{ \int_0^\infty \left[\left(\frac{\alpha}{2} + c \right) e_l^k - F(u_l^k) + F(u) \right] \exp(-\alpha t) dt \right\} g_l.$$

It is possible to suppose α to be large such that C^l belongs to $(\frac{\alpha}{4}, \alpha)$.

Lemma 1. *Choose g_l, $g_{l'}$ such that $\nabla g_l = \nabla g_{l'} = 0$ on $\Gamma_{l,l'}$ and $\frac{g_{l'}}{g_l} > 1$ on $\Gamma_{l,l'}$, for all l' in J_l. Φ_l^k is then a solution of the following equation*

$$\begin{cases} \mathfrak{L}_{lR}(\Phi_l^k) = 0, & \text{in } \Omega_l \times (0,\infty), \\ \beta_l \mathfrak{B}_{l,l'}(\Phi_l^k) = \mathfrak{B}_{l,l'}(\Phi_{l'}^{k-1}) & \text{on } \Gamma_{l,l'} \times (0,\infty), \forall l' \in J_l. \end{cases} \tag{4}$$

where $\beta_l = \frac{g_{l'}}{g_l}$ on $\Gamma_{l,l'}$, for all l' in J_l.

For all l in $\{1,I\}$, denote by $\tilde{\Omega}_l$ the open set $\Omega_l \backslash \overline{\cup_{l' \in J_l} \Omega_{l'}}$. For all l in I such that $\varphi_l^{k+1} = \varphi_{l'}^k$ on $\Gamma_{l,l'}$ for all l' in J_l, let φ_l^k and φ_l^{k+1} be functions in $H^1(\tilde{\Omega}_l)$ and $H^1(\Omega_l)$. Use the test functions φ_l^{k+1} and φ_l^k, and take the sum (with respect to l in $\{1,I\}$) of $\int_{\tilde{\Omega}_l} \mathfrak{L}_{lR}(\Phi_l^{k+1}) \varphi_l^{k+1}$ and $\int_{\tilde{\Omega}_l} \mathfrak{L}_{lR}(\Phi_l^k) \varphi_l^k$ to get

$$\begin{aligned} &-\sum_{l=1}^I \left\{ \int_{\tilde{\Omega}_l} C^l \Phi_l^k \varphi_l^k dx + \right. \\ &+ \int_{\tilde{\Omega}_l} \sum_{i,j=1}^n a_{i,j} \partial_i \Phi_l^k \partial_j \varphi_l^k dx + \sum_{i=1}^n \int_{\tilde{\Omega}_l} B_i^l \partial_i \Phi_l^k \varphi_l^k dx - \sum_{l' \in J_l} \int_{\Gamma_{l',l}} p_{l',l} \Phi_l^k \varphi_l^k d\sigma \\ &\left. + \int_{\tilde{\Omega}_l} \left\{ \int_0^\infty \left[\left(\frac{\alpha}{2} + c \right) e_l^k - F(u_l^k) + F(u) \right] \exp(-\alpha t) dt \right\} g_l \varphi_l^k dx \right\} \\ &= \sum_{l=1}^I \beta_l \left\{ \int_{\Omega_l} C^l \Phi_l^{k+1} \varphi_l^{k+1} dx + \right. \\ &+ \int_{\Omega_l} \sum_{i,j=1}^n a_{i,j} \partial_i \Phi_l^{k+1} \partial_j \varphi_l^{k+1} dx + \sum_{l' \in J_l} \int_{\Gamma_{l,l'}} p_{l,l'} \Phi_l^{k+1} \varphi_l^{k+1} d\sigma \\ &+ \int_{\Omega_l} \sum_{i=1}^n B_i^l \partial_i \Phi_l^{k+1} \varphi_l^{k+1} dx + \\ &\left. + \int_{\Omega_l} \left\{ \int_0^\infty \left[\left(\frac{\alpha}{2} + c \right) e_l^{k+1} - F(u_l^{k+1}) + F(u) \right] \exp(-\alpha t) dt \right\} g_l \varphi_l^{k+1} dx \right\}. \end{aligned} \tag{5}$$

In (5), choose φ_l^{k+1} to be Φ_l^{k+1}, then there exists φ_l^k, such that for all l' in J_l $\varphi_l^k = \varphi_{l'}^{k+1}$ on $\Gamma_{l,l'}$ and

$$||\varphi_l^k||_{H^1(\Omega_l)} \leq C \sum_{l' \in J_l} ||\varphi_{l'}^{k+1}||_{H^1(\Omega_{l'})} \text{ and } ||\varphi_l^k||_{L^2(\Omega_l)} \leq C \sum_{l' \in J_l} ||\varphi_{l'}^{k+1}||_{L^2(\Omega_{l'})},$$

where C is a positive constant.
The right hand side of (5) is then greater than or equal to

$$\begin{aligned} &\sum_{l=1}^{I} \beta_l \left\{ \int_{\Omega_l} \lambda |\nabla \Phi_l^{k+1}|^2 dx - \sum_{i=1}^{n} \int_{\Omega_l} ||B_i^l||_{L^\infty(\Omega_l)} \left| \partial_i \Phi_l^{k+1} \right| |\Phi_l^{k+1}| dx \right\}. \\ &\geq \sum_{l=1}^{I} \beta_l \left\{ \int_{\Omega_l} \frac{\lambda}{2} |\nabla \Phi_l^{k+1}|^2 dx + \frac{\alpha}{8} \int_{\Omega_l} |\Phi_l^{k+1}|^2 \right\}. \end{aligned} \tag{6}$$

Similarly, the left hand side of (5) is less than or equal to

$$\begin{aligned} &\sum_{l=1}^{I} \left\{ \int_{\tilde{\Omega}_l} \Lambda |\nabla \Phi_l^k| |\nabla \varphi_l^k| dx + \sum_{i=1}^{n} \int_{\tilde{\Omega}_l} ||B_i^l||_{L^\infty(\tilde{\Omega}_l)} \left| \partial_i \Phi_l^k \right| |\varphi_l^k| dx \right. \\ &\left. + \sum_{l' \in J_l} ||p_{l',l}||_{L^\infty(\Gamma_{l',l})} (||\Phi_l^k||^2_{H^1(\tilde{\Omega}_l)} + ||\varphi_l^k||^2_{H^1(\tilde{\Omega}_l)}) \right\} \\ &\leq \sum_{l=1}^{I} M_1 \left\{ \frac{1}{2} (||\nabla \Phi_l^k||^2_{L^2(\tilde{\Omega}_l)} + (\max_{i \in \{1,I\}} ||B_i^l||_{L^\infty(\tilde{\Omega}_l)})^2 ||\varphi_l^k||^2_{L^2(\tilde{\Omega}_l)}) \right. \\ &+ \int_{\tilde{\Omega}_l} 2\alpha |\Phi_l^k| |\varphi_l^k| dx + \sum_{l' \in J_l} \int_{\Gamma_{l',l}} p_{l',l} |\Phi_l^k| |\varphi_l^k| d\sigma \\ &\left. + \Lambda \left(||\nabla \Phi_l^k||^2_{L^2(\tilde{\Omega}_l)} + ||\nabla \varphi_l^k||^2_{L^2(\tilde{\Omega}_l)} \right) + \frac{\alpha}{2} ||\Phi_l^k||^2_{L^2(\tilde{\Omega}_l)} + \frac{\alpha}{2} ||\varphi_l^k||^2_{L^2(\tilde{\Omega}_l)} \right\}, \end{aligned} \tag{7}$$

where M_1 depends only on $\{\Omega_l\}_{l \in \{1,I\}}$ and the Eq. (3). Choose α such that $\alpha > (\max_{i \in \{1,I\}} ||B_i^l||_{L^\infty(\tilde{\Omega}_l)})^2$, there exists M_2 positive, depending only on $\{\Omega_l\}_{l \in \{1,I\}}$ and (3) such that the right hand side of (7) is dominated by

$$\begin{aligned} &\sum_{l=1}^{I} M_2 \left\{ \int_{\tilde{\Omega}_l} \left(\frac{\lambda}{2} |\nabla \Phi_l^k|^2 dx + \frac{\alpha}{8} |\Phi_l^k|^2 + \frac{\lambda}{2} |\nabla \Phi_l^{k+1}|^2 + \frac{\alpha}{8} |\Phi_l^{k+1}|^2 \right) dx \right\} \\ &\leq \sum_{l=1}^{I} M_2 \left(\frac{\lambda}{2} ||\nabla \Phi_l^k||^2_{L^2(\Omega_l)} + \frac{\alpha}{8} ||\Phi_l^k||^2_{L^2(\Omega_l)} + \frac{\lambda}{2} ||\nabla \Phi_l^{k+1}||^2_{L^2(\Omega_l)} + \frac{\alpha}{8} ||\Phi_l^{k+1}||^2_{L^2(\Omega_l)} \right). \end{aligned} \tag{8}$$

Define

$$E_k := \sum_{l=1}^{I} \left(\frac{\lambda}{2} ||\nabla \Phi_l^k||^2_{L^2(\Omega_l)} + \frac{\alpha}{8} ||\Phi_l^k||^2_{L^2(\Omega_l)} \right), \tag{9}$$

then (6), (7), and (8) imply

$$(\beta - M_2)E_{k+1} \leq M_2 E_k, \tag{10}$$

where $\beta = \min\{\beta_1, \ldots, \beta_I\}$.
Since M_2 depends only on $\{\Omega_l\}_{l\in\{1,I\}}$ and (3), β can be chosen such that

$$M_3 := \frac{M_2}{\beta - M_2} < 1.$$

We get

$$\begin{aligned} E_k &\leq M_3^k E_0 \\ &\leq M_3^k \sum_{l=1}^{I} \left(\frac{\lambda}{2} ||\nabla \Phi_l^0||^2_{L^2(\Omega_l)} + \frac{\alpha}{8} ||\Phi_l^0||^2_{L^2(\Omega_l)} \right). \end{aligned}$$

That deduces

$$||\Phi_l^k||^2_{L^2(\Omega_l)} \leq M_3^k \sum_{l=1}^{I} \left(\frac{4\lambda}{\alpha} ||\nabla \Phi_l^0||^2_{L^2(\Omega_l)} + ||\Phi_l^0||^2_{L^2(\Omega_l)} \right). \tag{11}$$

Since (11) still holds if M_3 and λ are fixed, and α is replaced by $y > \alpha$, then

$$\begin{aligned} &\sum_{l=1}^{I} \int_{\Omega_l} \left(\int_0^\infty e_l^k \exp(-yt) dt g_l \right)^2 dx \\ &\leq M_3^k \left[\frac{4\lambda}{y} \sum_{l=1}^{I} \int_{\Omega_l} \left(\int_0^\infty |\nabla e_l^0| \exp(-yt) dt \right)^2 g_l^2 dx \right. \\ &\quad + \frac{4\lambda}{y} \sum_{l=1}^{I} \int_{\Omega_l} \left(\int_0^\infty e_l^0 \exp(-yt) dt \right)^2 |\nabla g_l|^2 dx \\ &\quad \left. + \sum_{l=1}^{I} \int_{\Omega_l} \left(\int_0^\infty e_l^0 \exp(-yt) dt \right)^2 g_l^2 dx \right]. \end{aligned} \tag{12}$$

Let α' be a constant larger than or equal to α, (12) implies

$$\begin{aligned} &\sum_{l=1}^{I} \int_{\Omega_l} \int_{\alpha'}^{\alpha'+1} \left(\int_0^\infty e_l^k \exp(-yt) dt \right)^2 g_l^2 dy dx \\ &\leq M_3^k \left[\sum_{l=1}^{I} \int_{\Omega_l} \int_{\alpha'}^{\alpha'+1} \frac{4\lambda}{y} \left(\int_0^\infty |\nabla e_l^0| \exp(-yt) dt \right)^2 g_l^2 dy dx \right. \\ &\quad + \sum_{l=1}^{I} \int_{\Omega_l} \int_{\alpha'}^{\alpha'+1} \frac{4\lambda}{y} \left(\int_0^\infty e_l^0 \exp(-yt) dt \right)^2 |\nabla g_l|^2 dy dx \\ &\quad \left. + \sum_{l=1}^{I} \int_{\Omega_l} \int_{\alpha'}^{\alpha'+1} \left(\int_0^\infty e_l^0 \exp(-yt) dt \right)^2 g_l^2 dy dx \right]. \end{aligned} \tag{13}$$

Since u^0 belongs to $C_c^\infty(\overline{\Omega \times (0,\infty)})$, the right hand side of (13) is bounded by a constant $M_3^k M_4(\alpha)$. The fact that g_l is greater than 1 implies

$$\sum_{l=1}^{I} \int_{\Omega_l} \int_{\alpha'}^{\alpha'+1} \left(\int_0^{\infty} e_l^k \exp(-yt) dt \right)^2 dy dx \leq M_3^k M_4(\alpha). \tag{14}$$

Inequality (14) deduces

$$\lim_{k \to \infty} \sum_{l=1}^{I} \int_{\Omega_l} |e_l^k|_\alpha^2 dx = 0. \tag{15}$$

Acknowledgments The author would like to express his gratitude to his thesis advisor, Professor Laurence Halpern for her very kind help and support. He is also indebted to the editor for his kind help. The author has been supported by Grant MTM2011-29306-C02-00, MICINN, Spain, ERC Advanced Grant FP7-246775 NUMERIWAVES, and Grant PI2010-04 of the Basque Government.

Bibliography

[1] Lori Badea. On the Schwarz alternating method with more than two subdomains for nonlinear monotone problems. *SIAM J. Numer. Anal.*, 28(1):179–204, 1991. ISSN 0036-1429.

[2] Jean-David Benamou and Bruno Desprès. A domain decomposition method for the Helmholtz equation and related optimal control problems. *J. Comput. Phys.*, 136(1):68–82, 1997. ISSN 0021-9991.

[3] Martin J. Gander and Andrew M. Stuart. Space-time continuous analysis of waveform relaxation for the heat equation. *SIAM J. Sci. Comput.*, 19(6):2014–2031 (electronic), 1998. ISSN 1064-8275.

[4] Martin J. Gander and Hongkai Zhao. Overlapping Schwarz waveform relaxation for the heat equation in *n* dimensions. *BIT*, 42(4):779–795, 2002. ISSN 0006-3835.

[5] Eldar Giladi and Herbert B. Keller. Space-time domain decomposition for parabolic problems. *Numer. Math.*, 93(2):279–313, 2002. ISSN 0029-599X.

[6] Jung-Han Kimn. A convergence theory for an overlapping Schwarz algorithm using discontinuous iterates. *Numer. Math.*, 100(1):117–139, 2005. ISSN 0029-599X.

[7] P.-L. Lions. On the Schwarz alternating method. I. In *First International Symposium on Domain Decomposition Methods for Partial Differential Equations (Paris, 1987)*, pages 1–42. SIAM, Philadelphia, PA, 1988.

[8] P.-L. Lions. On the Schwarz alternating method. II. Stochastic interpretation and order properties. In *Domain decomposition methods (Los Angeles, CA, 1988)*, pages 47–70. SIAM, Philadelphia, PA, 1989.

[9] Sébastien Loisel and Daniel B. Szyld. On the geometric convergence of optimized Schwarz methods with applications to elliptic problems. *Numer. Math.*, 114(4):697–728, 2010. ISSN 0029-599X.

[10] S. H. Lui. On linear monotone iteration and Schwarz methods for nonlinear elliptic PDEs. *Numer. Math.*, 93(1):109–129, 2002. ISSN 0029-599X.

[11] Shiu-Hong Lui. On monotone and Schwarz alternating methods for nonlinear elliptic PDEs. *M2AN Math. Model. Numer. Anal.*, 35(1):1–15, 2001. ISSN 0764-583X.

[12] Frédéric Nataf and Francis Nier. Convergence of domain decomposition methods via semi-classical calculus. *Comm. Partial Differential Equations*, 23(5–6): 1007–1059, 1998. ISSN 0360-5302.

[13] Minh-Binh Tran. Convergence properties of overlapping Schwarz domain decomposition algorithms, submitted. http://arxiv.org/abs/1104.4294, 2012.

Part III

Contributed Presentations

FETI Methods for the Simulation of Biological Tissues

Christoph Augustin and Olaf Steinbach

Institute of Computational Mathematics, TU Graz, Steyrergasse 30, 8010 Graz, Austria,
c_augustin@gmx.net, o.steinbach@tugraz.at

Summary. In this paper we describe the application of finite element tearing and interconnecting methods for the simulation of biological tissues, as a particular application we consider the myocardium. As most other tissues, this material is characterized by anisotropic and nonlinear behavior.

1 Modeling Biological Tissues

In this paper we consider the numerical simulation of biological tissues, that can be described by the stationary equilibrium equations

$$\operatorname{div}\sigma(u,x)+f(x)=0 \quad \text{for } x\in\Omega\subset\mathbb{R}^3, \tag{1}$$

to find a displacement field u where we have to incorporate boundary conditions to describe the displacements or the boundary stresses on $\Gamma=\partial\Omega$.

In the case of biological tissues the material is assumed to be hyperelastic, i.e. we have to incorporate large deformations and a non-linear stress-strain relation. For the derivation of the constitutive equation we introduce the strain energy function $\Psi(\mathsf{C})$ which represents the elastic stored energy per unit reference volume. From this we obtain the constitutive equation as in [1]

$$\sigma=J^{-1}\mathsf{F}\frac{\partial\Psi(\mathsf{C})}{\partial\mathsf{C}}\mathsf{F}^\top,$$

where $J=\det\mathsf{F}$ is the Jacobian of the deformation gradient $\mathsf{F}=\nabla\varphi$, and $\mathsf{C}=\mathsf{F}^\top\mathsf{F}$ is the right Cauchy-Green tensor. In what follows we make use of the Rivlin-Ericksen representation theorem to find a representation of the strain energy function Ψ in terms of the principal invariants of $\mathsf{C}=\mathsf{F}^\top\mathsf{F}$.

The cardiac muscle, the so-called *myocardium*, is the most significant layer for the modeling of the elastic behavior of the heart wall. Muscle fibers are arranged in parallel, in different sheets within the tissue. Although this fiber type is predominant, we have also collagen that is arranged in a spatial network connecting the muscle

R. Bank et al. (eds.), *Domain Decomposition Methods in Science and Engineering XX*, 503
Lecture Notes in Computational Science and Engineering 91,
DOI 10.1007/978-3-642-35275-1_59,

fibers. We denote by $\mathbf{f}_0$ the *fiber axis* which is referred to as the main direction of the cardiac muscle fibers. The *sheet axis* $\mathbf{s}_0$ is defined to be perpendicular to $\mathbf{f}_0$ in the plane of the layer. This direction coincides with the collagen fiber orientation. As many other biological tissues we treat the myocardium as a nearly incompressible material. It shows a highly nonlinear and, due to the muscle and collagen fibers, an anisotropic behavior.

To capture the specifics of this fiber-reinforced composite, Holzapfel and Ogden proposed a strain-energy function Ψ that is decomposed into a volumetric, an isotropic and an anisotropic part, which consists of a transversely isotropic and an orthotropic response, see [7, 11],

$$\Psi(\mathsf{C}) = \Psi_{\rm vol}(J) + \Psi_{\rm iso}(\mathsf{C}) + \Psi_{\rm trans}(\mathsf{C},\mathbf{f}_0) + \Psi_{\rm trans}(\mathsf{C},\mathbf{s}_0) + \Psi_{\rm ortho}(\mathsf{C},\mathbf{f}_0,\mathbf{s}_0). \tag{2}$$

Following [11], we describe the volume changing part by

$$\Psi_{\rm vol}(J) = \frac{\kappa}{2}(\log J)^2. \tag{3}$$

The bulk modulus $\kappa > 0$ serves as a penalty parameter to enforce the (almost) incompressibility constraint. To model the isotropic ground substance we use a classical exponential model, see [2],

$$\Psi_{\rm iso}(\mathsf{C}) = \frac{a}{2b}\left\{\exp[b(J^{-2/3}I_1 - 3)] - 1\right\}, \tag{4}$$

where $a > 0$ is a stress-like and b is a dimensionless material parameter. $I_1 = \mathrm{tr}(\mathsf{C})$ is the first principal invariant of the right Cauchy-Green tensor C. In (2), $\Psi_{\rm trans}$ is associated with the deformations in direction of the fiber directions. Following [7] we describe the transversely isotropic response by using

$$\begin{aligned}
\Psi_{\rm trans}(\mathsf{C},\mathbf{f}_0) &= \frac{a_f}{2b_f}\left\{\exp[b_f(J^{-2/3}I_{4f} - 1)^2] - 1\right\}, \\
\Psi_{\rm trans}(\mathsf{C},\mathbf{s}_0) &= \frac{a_s}{2b_s}\left\{\exp[b_s(J^{-2/3}I_{4s} - 1)^2] - 1\right\},
\end{aligned} \tag{5}$$

with the invariants $I_{4f} := \mathbf{f}_0 \cdot (\mathsf{C}\mathbf{f}_0)$ and $I_{4s} := \mathbf{s}_0 \cdot (\mathsf{C}\mathbf{s}_0)$ and the material parameters a_f, b_f, a_s and a_f which are all assumed to be positive. It is worth to mention, that in this model the transversely isotropic responses $\Psi_{\rm trans}$ only contribute in the cases $I_{4f} > 1$, $I_{4s} > 1$, respectively. This corresponds to a stretch in a fiber direction, and this is explained by the wavy structure of the muscle and collagen fibers. In particular, the fibers are not able to support compressive stress. Moreover, the fibers are not active at low pressure, and the material behaves isotropically in this case. In contrast, at high pressure the collagen and muscle fibers straighten and then they govern the resistance to stretch of the material. This behavior of biological tissues was observed in experiments and this is fully covered by the myocardium model as described above. The stiffening effect at higher pressure also motivates the use of the exponential function in the anisotropic responses of the strain energy Ψ.

Finally a distinctive shear behavior motivates the inclusion of an orthotropic part in the strain energy function in terms of the invariant $I_{8fs} = \mathbf{f}_0 \cdot (\mathsf{C}\mathbf{s}_0)$

$$\Psi_{\text{ortho}}(\mathsf{C}) = \frac{a_{fs}}{2b_{fs}} \left\{ \exp(b_{fs} J^{-2/3} I_{8fs}^2) - 1 \right\}. \tag{6}$$

Here $a_{fs} > 0$ is a stress-like and $b_{fs} > 0$ a dimensionless material constant.

Note that the material parameters can be fitted to an experimentally observed response of the biological tissue. In the case of the myocardium, experimental data and, consequently, parameter sets are very rare. Following [7] and [11], we use the slightly adapted material parameters to be found in Table 1.

$\kappa = 3333.33$ kPa,	$a = 33.445$ kPa,	$b = 9.242$ (-),
$a_f = 18.535$ kPa,	$b_s = 10.446$ (-),	$b_f = 15.972$ (-),
$a_{fs} = 0.417$ kPa,	$a_s = 2.564$ kPa,	$b_{fs} = 11.602$ (-).

Table 1. Material parameters used in the numerical experiments [7, 11].

Note that similar models can also be used for the description of other biological materials, e.g., arteries, cf. [6, 8].

2 Finite Element Approximation

In this section we consider the variational formulation of the equilibrium equations (1) with Dirichlet boundary conditions $u = g_D$ on Γ_D, Neumann boundary conditions $t := \sigma(u)n = g_N$ on Γ_N, $\Gamma = \overline{\Gamma}_D \cup \overline{\Gamma}_N$, $\Gamma_D \cap \Gamma_N = \emptyset$, and n is the exterior normal vector of $\Gamma = \partial\Omega$. In particular we have to find $u \in [H^1(\Omega)]^3$, $u = g_D$ on Γ_D, such that

$$a(u,v) := \int_\Omega \sigma(u) : \mathsf{e}(v)\, dx = \int_\Omega f \cdot v\, dx + \int_{\Gamma_N} g_N \cdot v\, ds_x =: F(v) \tag{7}$$

is satisfied for all $v \in [H^1(\Omega)]^3$, $v = 0$ on Γ_D.

By introducing an admissible decomposition of the computational domain Ω into tetrahedra and by using piecewise quadratic basis functions φ_ℓ, the Galerkin finite element discretization of the variational formulation (7) results in a nonlinear system of algebraic equations, to find u_h satisfying an approximate Dirichlet boundary condition $u_h = Q_h g_D$ on Γ_D, and

$$K_\ell(u_h) = \int_\Omega \sigma(u_h) : \mathsf{e}(\varphi_\ell)\, dx = \int_\Omega f \cdot \varphi_\ell\, dx + \int_{\Gamma_N} g_N \cdot \varphi_\ell\, ds_x = F_\ell. \tag{8}$$

For the solution of the nonlinear system (8), i.e. of $G(u_h) := K(u_h) - F = 0$, we apply Newton's method to obtain the recursion

$$u_h^{k+1} = u_h^k + \Delta u_h^k, \quad \mathsf{G}_h'(u_h^k)\Delta u_h^k = -G(u_h^k),$$

or, by using the definition of $G(\cdot)$,

$$u_h^{k+1} = u_h^k + \Delta u_h^k, \quad \mathsf{K}_h'(u_h^k)\Delta u_h^k = -K(u_h^k). \tag{9}$$

For the computation of the linearized stiffness matrix $\mathsf{K}'_h(u_h^k)$ we need to evaluate the derivative of the nonlinear material model as described in the previous section. For a detailed presentation how to compute $\mathsf{K}'_h(u_h^k)$ in this particular case, see [5].

3 Finite Element Tearing and Interconnecting

For the parallel solution of (9) we will use a finite element tearing and interconnecting approach [4], see also [8, 14] and references given therein. For a bounded domain $\Omega \subset \mathbb{R}^3$ we introduce a non-overlapping domain decomposition

$$\overline{\Omega} = \bigcup_{i=1}^{p} \overline{\Omega}_i \quad \text{with } \Omega_i \cap \Omega_j = \emptyset \quad \text{for } i \neq j, \quad \Gamma_i = \partial\Omega_i. \tag{10}$$

The local interfaces are given by $\Gamma_{ij} := \Gamma_i \cap \Gamma_j$ for all $i < j$. The skeleton of the domain decomposition (10) is denoted as

$$\Gamma_C := \bigcup_{i=1}^{p} \Gamma_i = \Gamma \cup \bigcup_{i<j} \overline{\Gamma}_{ij}.$$

Instead of the global problem (1) we now consider local subproblems to find the local restrictions $u_i = u_{|\Omega_i}$ satisfying partial differential equations

$$\operatorname{div}(\sigma(u_i)) + f(x) = 0 \quad \text{for } x \in \Omega_i,$$

the Dirichlet and Neumann boundary conditions $u_i = g_D$ on $\Gamma_i \cap \Gamma_D$, $\sigma(u_i)n_i = g_N$ on $\Gamma_i \cap \Gamma_N$, and the transmission conditions $u_i = u_j$, $t_i + t_j = 0$ on Γ_{ij}, where $t_i = \sigma(u_i)n_i$ is the local boundary stress, and n_i is the exterior normal vector of the local subdomain boundary $\Gamma_i = \partial\Omega_i$. Note that the local stress tensors $\sigma(u_i)$ are defined locally by using the stress-strain function Ψ as introduced in Sect. 1, and by using localized parameters κ, k_1, k_2, c and fiber directions β_1, β_2. Hence, by reordering the degrees of freedom, the linearized system (9) can be written as

$$\begin{pmatrix} \mathsf{K}'_{11}(u_{1,h}^k) & & & \mathsf{K}'_{1C}(u_{1,h}^k)\mathsf{A}_1 \\ & \ddots & & \vdots \\ & & \mathsf{K}'_{pp}(u_{p,h}^k) & \mathsf{K}'_{pC}(u_{p,h}^k)\mathsf{A}_p \\ \mathsf{A}_1^\top \mathsf{K}'_{C1}(u_{1,h}^k) & \cdots & \mathsf{A}_p^\top \mathsf{K}'_{Cp}(u_{p,h}^k) & \sum_{i=1}^{p} \mathsf{A}_i^\top \mathsf{K}'_{CC}(u_{i,h}^k)\mathsf{A}_i \end{pmatrix} \begin{pmatrix} \Delta\mathbf{u}_{1,I}^k \\ \vdots \\ \Delta\mathbf{u}_{p,I}^k \\ \Delta\mathbf{u}_C^k \end{pmatrix} = - \begin{pmatrix} \mathsf{K}_1(u_{1,h}^k) \\ \vdots \\ \mathsf{K}_p(u_{p,h}^k) \\ \sum_{i=1}^{p} \mathsf{A}_i^\top \mathsf{K}_C(u_{i,h}^k) \end{pmatrix},$$

where the increments $\Delta\mathbf{u}_{i,I}^k$ correspond to the local degrees of freedom within the subdomain Ω_i, and $\Delta\mathbf{u}_C^k$ is related to all global degrees of freedom on the coupling boundary Γ_C. By introducing the tearing

$$\mathbf{w}_i = \begin{pmatrix} \Delta\mathbf{u}_{i,I}^k \\ \mathsf{A}_i \Delta\mathbf{u}_C^k \end{pmatrix}, \quad \mathsf{K}'_i = \begin{pmatrix} \mathsf{K}'_{ii}(u_{i,h}^k) & \mathsf{K}'_{iC}(u_{i,h}^k) \\ \mathsf{K}'_{Ci}(u_{i,h}^k) & \mathsf{K}'_{CC}(u_{i,h}^k) \end{pmatrix}, \quad \mathbf{f}_i = - \begin{pmatrix} \mathsf{K}_i(u_{i,h}^k) \\ \mathsf{K}_C(u_{i,h}^k) \end{pmatrix},$$

by applying the interconnecting $\sum_{i=1}^{p} \mathsf{B}_i \mathbf{w}_i = \mathbf{0}$, and by using discrete Lagrange multipliers, we finally have to solve the system

$$\begin{pmatrix} \mathsf{K}_1' & & & \mathsf{B}_1^\top \\ & \ddots & & \vdots \\ & & \mathsf{K}_p' & \mathsf{B}_p^\top \\ \mathsf{B}_1 & \dots & \mathsf{B}_p & \end{pmatrix} \begin{pmatrix} \mathbf{w}_1 \\ \vdots \\ \mathbf{w}_p \\ \lambda \end{pmatrix} = \begin{pmatrix} \mathbf{f}_1 \\ \vdots \\ \mathbf{f}_p \\ \mathbf{0} \end{pmatrix}. \tag{11}$$

For the solution of the linear system (11) we follow the standard approach of tearing and interconnecting methods. In the case of a floating subdomain Ω_i, i.e. $\Gamma_i \cap \Gamma_D = \emptyset$, the local matrices K_i' are not invertible. Hence we introduce the Moore-Penrose pseudo inverse $\mathsf{K}_i^\dagger$ to represent the local solutions as

$$\mathbf{w}_i = \mathsf{K}_i^\dagger (\mathbf{f}_i - \mathsf{B}_i^\top \lambda) + \sum_{k=1}^{6} \gamma_{k,i} \mathbf{v}_{k,i}, \tag{12}$$

where $\mathbf{v}_{k,i} \in \ker \mathsf{K}_i'$ correspond to the rigid body motions of elasticity. Note that in this case we also require the solvability conditions

$$(\mathbf{f}_i - \mathsf{B}_i^\top \lambda, \mathbf{v}_{k,i}) = 0 \quad \text{for } i = 1, \dots, 6.$$

In the case of a non-floating subdomain, i.e. $\ker \mathsf{K}_i = \emptyset$, we may set $\mathsf{K}_i^\dagger = \mathsf{K}_i^{-1}$. As in [10] we may also consider an all-floating approach where also Dirichlet boundary conditions are incorporated by using discrete Lagrange multipliers.

In general, we consider the Schur complement system of (11) to obtain

$$\sum_{i=1}^{p} \mathsf{B}_i \mathsf{K}_i^\dagger \mathsf{B}_i^\top \lambda - \sum_{i=1}^{p} \sum_{k=1}^{6} \gamma_{k,i} \mathsf{B}_i \mathbf{v}_{k,i} = \sum_{i=1}^{p} \mathsf{B}_i \mathsf{K}_i^\dagger \mathbf{f}_i, \quad (\mathbf{f}_i - \mathsf{B}_i^\top \lambda, \mathbf{v}_{k,i}) = 0,$$

which can be written as

$$\begin{pmatrix} \mathsf{F} & -\mathsf{G} \\ \mathsf{G}^\top & \end{pmatrix} \begin{pmatrix} \lambda \\ \gamma \end{pmatrix} = \begin{pmatrix} \mathbf{d} \\ \mathbf{e} \end{pmatrix} \tag{13}$$

with

$$\mathsf{F} = \sum_{i=1}^{p} \mathsf{B}_i \mathsf{K}_i^\dagger \mathsf{B}_i^\top, \ \mathsf{G} = \sum_{i=1}^{p} \sum_{k=1}^{6} \mathsf{B}_i \mathbf{v}_{k,i}, \ \mathbf{d} = \sum_{i=1}^{p} \mathsf{B}_i \mathsf{K}_i^\dagger \mathbf{f}_i, \ e_{k,i} = (\mathbf{f}_i, \mathbf{v}_{k,i}).$$

For the solution of the linear system (13) we use the projection $\mathsf{P}^\top := \mathsf{I} - \mathsf{G}(\mathsf{G}^\top \mathsf{G})^{-1} \mathsf{G}^\top$ and it remains to consider the projected system

$$\mathsf{P}^\top \mathsf{F} \lambda = \mathsf{P}^\top \mathbf{d} \tag{14}$$

which can be solved by using a parallel GMRES method with suitable preconditioning. Note that the initial approximate solution λ^0 satisfies the compatibility condition $\mathsf{G}^\top \lambda^0 = \mathbf{e}$. In a post processing we finally recover $\gamma = (\mathsf{G}^\top \mathsf{G})^{-1} \mathsf{G}^\top (\mathsf{F}\lambda - \mathbf{d})$, and subsequently the desired solution (12).

Following [3] we are going to apply either the lumped preconditioner

$$\mathsf{PM}^{-1} := \sum_{i=1}^{p} \mathsf{B}_i \mathsf{K}_i' \mathsf{B}_i^\top, \tag{15}$$

or the Dirichlet preconditioner

$$\mathsf{PM}^{-1} := \sum_{i=1}^{p} \mathsf{B}_i \begin{pmatrix} 0 & 0 \\ 0 & \mathsf{S}_i \end{pmatrix} \mathsf{B}_i^\top, \tag{16}$$

where

$$\mathsf{S}_i = \mathsf{K}_{CC}'(u_{i,h}^k) - \mathsf{K}_{Ci}'(u_{i,h}^k)\mathsf{K}_{ii}'^{-1}(u_{i,h}^k)\mathsf{K}_{iC}'(u_{i,h}^k)$$

is the Schur complement of the local finite element matrix K_i'. Alternatively, one may also use the scaled hypersingular boundary integral operator preconditioner as proposed in [9].

4 Numerical Results

In this section we present some examples to show the applicability of the FETI approach for the simulation of the myocardium, see Fig. 3. We consider a mesh of the left and the right ventricle of a rabbit heart with given fiber and sheet directions, see Fig. 1, which is decomposed in 480 subdomains, see Fig. 2. To describe the anisotropic and nonlinear cardiac tissue, we use the material model (2) with the parameters given in Table 1. Dirichlet boundary conditions are imposed on the top of the myocardium mesh. The interior wall of the right ventricle is exposed to the pressure of 1 mmHg which is modeled with Neumann boundary conditions. Although this pressure is rather low, the material model as used is orthotropic. To simulate a higher pressure, an appropriate time stepping scheme has to be used. However, this does not affect the number of local iterations significantly. The local Moore Penrose pseudo inverse matrices are realized with a sparsity preserving regularization and the direct solver package Pardiso [12, 13]. The global nonlinear finite element system with 12.188.296 degrees of freedom is solved by a Newton scheme, where the FETI approach is used in each Newton step. For this specific example the Newton scheme needed six iterations. Due to the non-uniformity of the subdomains the efficiency of a global preconditioner becomes more important. We consider both the classical FETI approach, as well as the all–floating formulation. Besides no preconditioning we use the simple lumped preconditioner (15) and the Dirichlet preconditioner (16). It turns out that the number of iterations for the all–floating formulation is approximately half the number of iterations for the standard approach. Moreover, the Dirichlet preconditioner within the all–floating formulation requires only 108 iterations, with a computing time of approximately 5 min. All computations were done at the Vienna Scientific Cluster (VSC2).

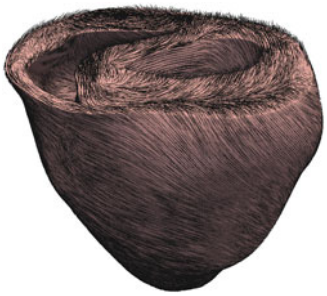


Fig. 1. *Left* and *right* ventricle of the rabbit heart. Mesh consists of 3.073.529 tetrahedrons and 547.680 vertices. *Black lines* indicate fiber directions $\mathbf{f}_0$. Point of view is from above showing the interior of the *left* and *right* ventricle

preconditioner	iterations
classical FETI	
none	941
lumped, (15)	916
Dirichlet, (16)	215
all-floating FETI	
none	535
lumped, (15)	401
Dirichlet, (16)	108

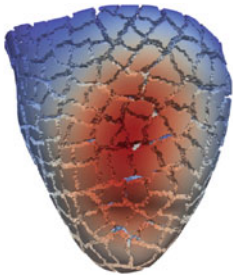


Fig. 2. The picture shows the displacement field of the rabbit heart with pressure applied in the *right* ventriculum. Point of view is from below showing the apex of the heart at the *bottom*. In the table the iteration numbers of the global GMRES method for different preconditioners are given

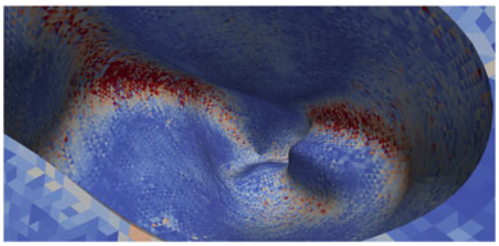


Fig. 3. Von Mises stress in the *right* ventricle. Point of view is from above looking inside the *right* ventricle

Acknowledgments This work was supported by the Austrian Science Fund (FWF) and by the TU Graz within the SFB Mathematical Optimization and Applications in Biomedical Sciences. The authors would like to thank G. A. Holzapfel, G. Of, G. Plank, and C. Pechstein for the fruitful cooperation and many helpful discussions. We also thank the referees for their helpful remarks and suggestions.

Bibliography

[1] P. G. Ciarlet. *Mathematical elasticity. Vol. I*, volume 20 of *Studies in Mathematics and its Applications*. North-Holland, Amsterdam, 1988.

[2] H. Demiray. A note on the elasticity of soft biological tissues. *J. Biomech.*, 5:309–311, 1972.

[3] C. Farhat, J. Mandel, and F.-X. Roux. Optimal convergence properties of the FETI domain decomposition method. *Comput. Methods Appl. Mech. Engrg.*, 115:365–385, 1994.

[4] C. Farhat and F.-X. Roux. A method of finite element tearing and interconnecting and its parallel solution algorithm. *Internat. J. Numer. Methods Engrg.*, 32:1205–1227, 1991.

[5] G. A. Holzapfel. Structural and numerical models for the (visco)elastic response of arterial walls with residual stresses. In G. A. Holzapfel and R. W. Ogden, editors, *Biomechanics of Soft Tissue in Cardiovascular Systems*. Springer, Wien, New York, 2003.

[6] G. A. Holzapfel, T. C. Gasser, and R. W. Ogden. A new constitutive framework for arterial wall mechanics and a comperative study of material models. *J. Elasticity*, 61:1–48, 2000.

[7] G. A. Holzapfel and R. W. Ogden. Constitutive modelling of passive myocardium: a structurally based framework for material characterization. *Phil. Trans. Math. Phys. Eng. Sci.*, 367:3445–3475, 2009.

[8] A. Klawonn and O. Rheinbach. Highly scalable parallel domain decomposition methods with an application to biomechanics. *ZAMM Z. Angew. Math. Mech.*, 90:5–32, 2010.

[9] U. Langer and O. Steinbach. Boundary element tearing and interconnecting methods. *Computing*, 71:205–228, 2003.

[10] G. Of and O. Steinbach. The all–floating boundary element tearing and interconnecting method. *J. Numer. Math.*, 7:277–298, 2009.

[11] T. S. E. Eriksson A. J. Prassl, G. Plank, and G. A. Holzapfel. Modelling the electromechanically coupled orthotropic structure of myocardium. *Submitted*.

[12] O. Schenk, M. Bollhöfer, and R. A. Römer. On large scale diagonalization techniques for the Anderson model of localization. *SIAM Review*, 50(1):91–112, 2008. SIGEST Paper.

[13] O. Schenk, A. Wächter, and M. Hagemann. Matching-based preprocessing algorithms to the solution of saddle-point problems in large-scale nonconvex interior-point optimization. *Comput. Optim. Appl.*, 36(2–3):321–341, 2007.

[14] A. Toselli and O. B. Widlund. *Domain Decomposition Methods – Algorithms and Theory*. Springer, Berlin, Heidelberg, 2005.

Fast Summation Techniques for Sparse Shape Functions in Tetrahedral *hp*-FEM

Sven Beuchler[1], Veronika Pillwein[2], and Sabine Zaglmayr[3]

[1] Institute for Numerical Simulation, University Bonn, Wegelerstraße 6, 53115 Bonn, Germany beuchler@ins.uni-bonn.de
[2] RISC, University Linz, Altenberger Straße 69, 4040 Linz, Austria pillwein@risc.uni-linz.ac.at
[3] CST - Computer Simulation Technology AG, Darmstadt, Germany., sabine.zaglmayr@cst.com

Summary. This paper considers the *hp*-finite element discretization of an elliptic boundary value problem using tetrahedral elements. The discretization uses a polynomial basis in which the number of nonzero entries per row is bounded independently of the polynomial degree. The authors present an algorithm which computes the nonzero entries of the stiffness matrix in optimal complexity. The algorithm is based on sum factorization and makes use of the nonzero pattern of the stiffness matrix.

1 Introduction

hp-finite element methods (*hp*-FEM), see e.g. [6, 9], have become very popular for the approximation of solutions of boundary value problems with more regularity. In order to obtain the approximate finite element solution numerically stable and fast, the functions have to be chosen properly in *hp*-FEM. For quadrilateral and hexahedral elements, tensor products of integrated Legendre polynomials are the prefered basis functions, see [2]. For triangular and tetrahedral elements, the element can be considered as collapsed quadrilateral or hexahedron. This allows us to use tensor product functions. In order to obtain sparsity in the element matrices and a moderate increase of the condition number, integrated Jacobi polynomials can be used, see [3, 5, 7]. Then, it has been shown that the element stiffness and mass matrix have a bounded number of nonzero entries per row, see [3–5] which results in a total number of $\mathcal{O}(p^d)$, $d = 2, 3$, nonzero entries in two and three space dimensions, respectively. However, the explicit computation of the nonzero entries is very involved.

This paper presents an algorithm which computes the element stiffness and mass matrices in $\mathcal{O}(p^3)$ operations in two and three space dimensions. The algorithm combines ideas based on sum factorization, [8], with the sparsity pattern of the matrices. One other important ingredient is the fast evaluation of the Jacobi polynomials.

The outline of this paper is as follows. Section 2 defines H^1-conforming, i.e. globally continuous piecewise polynomials, basis functions on the tetrahedron. The

R. Bank et al. (eds.), *Domain Decomposition Methods in Science and Engineering XX*, Lecture Notes in Computational Science and Engineering 91,
DOI 10.1007/978-3-642-35275-1_60,

sum factorization algorithm is presented in Sect. 3. Section 4 is devoted to the evaluation of the Jacobi polynomials. The complexity of the algorithm is estimated in Sect. 5.

2 Definition of the Basis Functions

For the definition of our basis functions Jacobi polynomials are required. Let

$$p_n^{\alpha}(x) = \frac{1}{2^n n!(1-x)^{\alpha}} \frac{\mathrm{d}^n}{\mathrm{d}x^n} \left((1-x)^{\alpha}(x^2-1)^n \right) \quad n \in \mathbb{N}_0,\ \alpha, \beta > -1 \tag{1}$$

be the nth Jacobi polynomial with respect to the weight function $(1-x)^{\alpha}$. The function p_n^{α} is a polynomial of degree n, i.e., $p_n^{\alpha} \in \mathbb{P}_n((-1,1))$, where $\mathbb{P}_n(I)$ is the space of all polynomials of degree n on the interval I. In the special case $\alpha = 0$, the functions $p_n^0(x)$ are called Legendre polynomials. Moreover, let

$$\hat{p}_n^{\alpha}(x) = \int_{-1}^{x} p_{n-1}^{\alpha}(y)\,\mathrm{d}y \quad n \geq 1, \quad \hat{p}_0^{\alpha}(x) = 1 \tag{2}$$

be the nth integrated Jacobi polynomial. Several relations are known between the different families of Jacobi polynomials, see e.g. [1]. In this paper, the relations

$$p_n^{\alpha-1}(x) = \frac{1}{\alpha+2n}\left[(\alpha+n)p_n^{\alpha}(x) - n p_{n-1}^{\alpha}(x)\right], \tag{3}$$

$$\begin{aligned} \hat{p}_{n+1}^{\alpha}(x) = {} & \frac{2n+\alpha-1}{(2n+2)(n+\alpha)(2n+\alpha-2)} \\ & \times \left((2n+\alpha-2)(2n+\alpha)x + \alpha(\alpha-2)\right)\hat{p}_n^{\alpha}(x) \\ & - \frac{(n-1)(n+\alpha-2)(2n+\alpha)}{(n+1)(n+\alpha)(2n+\alpha-2)}\hat{p}_{n-1}^{\alpha}(x), \quad n \geq 1. \end{aligned} \tag{4}$$

are required.

Let $\hat{\triangle}$ be the reference tetrahedron with the four vertices at $(-1,-1,-1)$, $(1,-1,-1)$, $(0,1,-1)$ and $(0,0,1)$. On this element, the interior bubble functions

$$\phi_{ijk}(x,y,z) = u_i(x,y,z)v_{ij}(y,z)w_{ijk}(z), \quad i \geq 2,\ j,k \geq 1, i+j+k \leq p \tag{5}$$

are proposed for H^1 elliptic problems in [3, (29)], where the auxiliary functions are

$$\begin{aligned} u_i(x,y,z) &= \hat{p}_i^0\left(\frac{4x}{1-2y-z}\right)\left(\frac{1-2y-z}{4}\right)^i, \\ v_{ij}(y,z) &= \hat{p}_j^{2i-1}\left(\frac{2y}{1-z}\right)\left(\frac{1-z}{2}\right)^j, \\ w_{ijk}(z) &= \hat{p}_k^{2i+2j-2}(z). \end{aligned}$$

In addition, there are vertex, face and edge based basis functions which can be regarded as special cases of the above functions (5) for limiting cases of the indices i, j and k, see [3] for more details.

Then, the element stiffness matrix for the Laplacian on the reference element $\hat{\triangle}$ with respect to the interior bubbles reads as

$$\mathscr{K} = \left[\int_{\hat{\triangle}} \nabla \phi_{ijk}(x,y,z) \cdot \nabla \phi_{i'j'k'}(x,y,z) \, \mathrm{d}(x,y,z) \right]_{i,j,k \le p, i'+j'+k' \le p}. \tag{6}$$

The transformation to the unit cube $(-1,1)^3$ (Duffy trick) and the evaluation of the nabla operation results in the integration of 21 different summands. More precisely,

$$\mathscr{K} = \sum_{m=1}^{21} \kappa_m \hat{\mathscr{I}}^{(m)}$$

with known numbers $\kappa_m \in \mathbb{R}$ (see Table 2) and

$$\begin{aligned} \hat{\mathscr{I}}^{(m)} = \Bigg[& \int_{-1}^{1} p_{x,1}(x) p_{x,2}(x) \, dx \\ & \times \int_{-1}^{1} \left(\frac{1-y}{2} \right)^{\gamma_y} p_{y,1}(y) p_{y,2}(y) \, \mathrm{d}y \\ & \times \int_{-1}^{1} \left(\frac{1-z}{2} \right)^{\gamma_z} p_{z,1}(z) p_{z,2}(z) \, \mathrm{d}z \Bigg]_{i+j+k<p; i'+j'+k'<p}. \end{aligned}$$

The structure of the functions and coefficients is displayed in Table 1.

One summand is the term

$$\hat{\mathscr{I}}^{(6)} = \left(m_{ijk,i'j'k'} \right)_{i+j+k \le p, i'+j'+k' \le p} \tag{7}$$

which corresponds (before the Duffy trick) to

$$\begin{aligned} m_{ijk,i'j'k'} = & \int_{\hat{\triangle}} \hat{p}_i^0 \left(\frac{4x}{1-2y-z} \right) \hat{p}_{i'}^0 \left(\frac{4x}{1-2y-z} \right) \left(\frac{1-2y-z}{4} \right)^{i+i'} \\ & \times \hat{p}_j^{2i-1} \left(\frac{2y}{1-z} \right) \hat{p}_{j'}^{2i'-1} \left(\frac{2y}{1-z} \right) \left(\frac{1-z}{2} \right)^{j+j'} \\ & \times p_{k-1}^{2i+2j-2}(z) p_{k'-1}^{2i'+2j'-2}(z) \, \mathrm{d}(x,y,z). \end{aligned}$$

The Duffy transformation applied to (7) gives

$$\begin{aligned} m_{ijk,i'j'k'} = & \int_{-1}^{1} \hat{p}_i^0(x) \hat{p}_{i'}^0(x) \, dx \int_{-1}^{1} \left(\frac{1-y}{2} \right)^{i+i'+1} \hat{p}_{j'}^{2i'-1}(y) \hat{p}_j^{2i-1}(y) \, \mathrm{d}y \\ & \times \int_{-1}^{1} \left(\frac{1-z}{2} \right)^{i+j+i'+j'+2} p_{k-1}^{2i+2j-2}(z) p_{k'-1}^{2i'+2j'-2}(z) \, \mathrm{d}z. \end{aligned} \tag{8}$$

It has been shown in [3], this matrix has the sparsity pattern

$$m_{ijk,i'j'k'} = 0 \quad \text{if } (i,j,k,i',j',k') \in \mathfrak{S}_{ref}^p(ijk,i'j'k') \tag{9}$$

	$p_{x,1}$	$p_{x,2}$	γ_y	$p_{y,1}$	$p_{y,2}$	γ_z	$p_{z,1}$	$p_{z,2}$
$\hat{\mathscr{I}}^{(1)}$	p^0_{i-1}	$p^0_{i'-1}$	$i+i'-1$	$\hat{p}^{2i-1}_j$	$\hat{p}^{2i'-1}_{j'}$	$\beta+\beta'$	$\hat{p}^{-2+2\beta}_k$	$\hat{p}^{-2+2\beta'}_{k'}$
$\hat{\mathscr{I}}^{(2)}$	$\hat{p}^0_i$	$\hat{p}^0_{i'}$	$i+i'+1$	p^{2i-1}_{j-1}	$p^{2i'-1}_{j'-1}$	$\beta+\beta'$	$\hat{p}^{-2+2\beta}_k$	$\hat{p}^{-2+2\beta'}_{k'}$
$\hat{\mathscr{I}}^{(3)}$	p^0_{i-2}	$\hat{p}^0_{i'}$	$i+i'$	$\hat{p}^{2i-1}_j$	$p^{2i'-1}_{j'-1}$	$\beta+\beta'$	$\hat{p}^{-2+2\beta}_k$	$\hat{p}^{-2+2\beta'}_{k'}$
$\hat{\mathscr{I}}^{(4)}$	$\hat{p}^0_i$	$p^0_{i'-2}$	$i+i'$	p^{2i-1}_{j-1}	$\hat{p}^{2i'-1}_{j'}$	$\beta+\beta'$	$\hat{p}^{-2+2\beta}_k$	$\hat{p}^{-2+2\beta'}_{k'}$
$\hat{\mathscr{I}}^{(5)}$	p^0_{i-2}	$p^0_{i'-2}$	$i+i'-1$	$\hat{p}^{2i-1}_j$	$\hat{p}^{2i'-1}_{j'}$	$\beta+\beta'$	$\hat{p}^{-2+2\beta}_k$	$\hat{p}^{-2+2\beta'}_{k'}$
$\hat{\mathscr{I}}^{(6)}$	$\hat{p}^0_i$	$\hat{p}^0_{i'}$	$i+i'+1$	$\hat{p}^{2i-1}_j$	$\hat{p}^{2i'-1}_{j'}$	$\beta+\beta'+2$	$p^{-2+2\beta}_{k-1}$	$p^{-2+2\beta'}_{k'-1}$
$\hat{\mathscr{I}}^{(7)}$	$\hat{p}^0_i$	$\hat{p}^0_{i'}$	$i+i'+1$	p^{2i-1}_{j-2}	$\hat{p}^{2i'-1}_{j'}$	$\beta+\beta'+1$	$\hat{p}^{-2+2\beta}_k$	$p^{-2+2\beta'}_{k'-1}$
$\hat{\mathscr{I}}^{(8)}$	$\hat{p}^0_i$	$\hat{p}^0_{i'}$	$i+i'+1$	p^{2i-1}_{j-1}	$\hat{p}^{2i'-1}_{j'}$	$\beta+\beta'+1$	$\hat{p}^{-2+2\beta}_k$	$p^{-2+2\beta'}_{k'-1}$
$\hat{\mathscr{I}}^{(9)}$	p^0_{i-2}	$\hat{p}^0_{i'}$	$i+i'$	$\hat{p}^{2i-1}_j$	$\hat{p}^{2i'-1}_{j'}$	$\beta+\beta'+1$	$\hat{p}^{-2+2\beta}_k$	$p^{-2+2\beta'}_{k'-1}$
$\hat{\mathscr{I}}^{(10)}$	$\hat{p}^0_i$	$\hat{p}^0_{i'}$	$i+i'+1$	$\hat{p}^{2i-1}_j$	$p^{2i'-1}_{j'-2}$	$\beta+\beta'+1$	$p^{-2+2\beta}_{k-1}$	$\hat{p}^{-2+2\beta'}_{k'}$
$\hat{\mathscr{I}}^{(11)}$	$\hat{p}^0_i$	$\hat{p}^0_{i'}$	$i+i'+1$	$\hat{p}^{2i-1}_j$	$p^{2i'-1}_{j'-1}$	$\beta+\beta'+1$	$p^{-2+2\beta}_{k-1}$	$\hat{p}^{-2+2\beta'}_{k'}$
$\hat{\mathscr{I}}^{(12)}$	$\hat{p}^0_i$	$\hat{p}^0_{i'}$	$i+i'+1$	p^{2i-1}_{j-2}	$p^{2i'-1}_{j'-2}$	$\beta+\beta'$	$\hat{p}^{-2+2\beta}_k$	$\hat{p}^{-2+2\beta'}_{k'}$
$\hat{\mathscr{I}}^{(13)}$	$\hat{p}^0_i$	$\hat{p}^0_{i'}$	$i+i'+1$	p^{2i-1}_{j-1}	$p^{2i'-1}_{j'-2}$	$\beta+\beta'$	$\hat{p}^{-2+2\beta}_k$	$\hat{p}^{-2+2\beta'}_{k'}$
$\hat{\mathscr{I}}^{(14)}$	$\hat{p}^0_i$	$\hat{p}^0_{i'}$	$i+i'+1$	p^{2i-1}_{j-2}	$p^{2i'-1}_{j'-1}$	$\beta+\beta'$	$\hat{p}^{-2+2\beta}_k$	$\hat{p}^{-2+2\beta'}_{k'}$
$\hat{\mathscr{I}}^{(15)}$	$\hat{p}^0_i$	$\hat{p}^0_{i'}$	$i+i'+1$	p^{2i-1}_{j-1}	$p^{2i'-1}_{j'-1}$	$\beta+\beta'$	$\hat{p}^{-2+2\beta}_k$	$\hat{p}^{-2+2\beta'}_{k'}$
$\hat{\mathscr{I}}^{(16)}$	p^0_{i-2}	$\hat{p}^0_{i'}$	$i+i'$	$\hat{p}^{2i-1}_j$	$p^{2i'-1}_{j'-2}$	$\beta+\beta'$	$\hat{p}^{-2+2\beta}_k$	$\hat{p}^{-2+2\beta'}_{k'}$
$\hat{\mathscr{I}}^{(17)}$	p^0_{i-2}	$\hat{p}^0_{i'}$	$i+i'$	$\hat{p}^{2i-1}_j$	$p^{2i'-1}_{j'-1}$	$\beta+\beta'$	$\hat{p}^{-2+2\beta}_k$	$\hat{p}^{-2+2\beta'}_{k'}$
$\hat{\mathscr{I}}^{(18)}$	$\hat{p}^0_i$	$p^0_{i'-2}$	$i+i'$	$\hat{p}^{2i-1}_j$	$\hat{p}^{2i'-1}_{j'}$	$\beta+\beta'+1$	$p^{-2+2\beta}_{k-1}$	$\hat{p}^{-2+2\beta'}_{k'}$
$\hat{\mathscr{I}}^{(19)}$	$\hat{p}^0_i$	$p^0_{i'-2}$	$i+i'$	p^{2i-1}_{j-2}	$\hat{p}^{2i'-1}_{j'}$	$\beta+\beta'$	$\hat{p}^{-2+2\beta}_k$	$\hat{p}^{-2+2\beta'}_{k'}$
$\hat{\mathscr{I}}^{(20)}$	$\hat{p}^0_i$	$p^0_{i'-2}$	$i+i'$	p^{2i-1}_{j-1}	$\hat{p}^{2i'-1}_{j'}$	$\beta+\beta'$	$\hat{p}^{-2+2\beta}_k$	$\hat{p}^{-2+2\beta'}_{k'}$
$\hat{\mathscr{I}}^{(21)}$	p^0_{i-2}	$p^0_{i'-2}$	$i+i'-1$	$\hat{p}^{2i-1}_j$	$\hat{p}^{2i'-1}_{j'}$	$\beta+\beta'$	$\hat{p}^{-2+2\beta}_k$	$\hat{p}^{-2+2\beta'}_{k'}$

Table 1. Integrands for $\mathscr{K}$, where $\beta = i+j$, $\beta' = i'+j'$

where

$$\mathfrak{S}^p_{ref}(ijk,i'j'k') = \{i+j+k \le p, i'+j'+k' \le p, |i-i'| \notin \{0,2\}$$
$$\vee \quad |i-i'+j-j'| > 4 \qquad \vee \quad |i-i'+j-j'+k-k'| > 4\}$$

cf. [3, Theorem 3.3]. In the following the more general case

$$\mathfrak{S}^p(ijk,i'j'k') = \{i+j+k \le p, i'+j'+k' \le p, |i-i'| > 2 \tag{10}$$
$$\vee \quad |i-i'+j-j'| > 4 \qquad \vee \quad |i-i'+j-j'+k-k'| > 4\}$$

is considered, e.g. the orthogonalities for $|i-i'| = 1$ are not assumed.

All 21 integrals give rise to a similar band structure as detailed above for $\hat{\mathscr{I}}^{(6)}$ and can thus be treated in the same way as explained below for the particular case

m	κ_m
$1,6,9,19$	1
$5,21$	$\frac{5}{4}$
$4,8,20$	$c_1(i,j)$
$7,19$	$c_2(i,j)$
$3,11,17$	$c_1(i',j')$
$2,15$	$c_1(i,j)c_1(i',j')$
13	$c_1(i,j)c_2(i',j')$
$10,16$	$c_2(i',j')$
14	$c_1(i',j')c_2(i,j)$
21	$c_2(i,j)c_2(i',j')$

Table 2. Coefficients κ_m for $\mathscr{K}$, where $c_1(i,j) = -\frac{1}{2}\frac{2i-1}{2i+2j-3}$ and $c_2(i,j) = \frac{j-1}{2i+2j-3}$.

of $\hat{\mathscr{I}}^{(6)}$. The only difference are shifts in the weights α of the Jacobi polynomials or changes of the weight functions.

3 Sum Factorization

In this section, we present an algorithm for the fast numerical generation of the local element matrices (6) for tetrahedra. The methods are based on fast summation techniques presented in [7, 8] and are carried out in detail for the example of the matrix $\hat{\mathscr{I}}^{(6)}$ (8).

All one dimensional integrals in (8) are computed numerically by a Gaussian quadrature rule with points x_k, $k = 1,\dots,p+1$ and corresponding weights ω_k. The points and weights are chosen such that

$$\int_{-1}^{1} f(x)\,dx = \sum_{l=1}^{p+1} \omega_l f(x_l) \quad \forall f \in \mathscr{P}_{2p}. \tag{11}$$

Since only polynomials of maximal degree $2p$ are integrated in (8), these integrals are evaluated exactly. Therefore, we have to compute

$$
\begin{aligned}
m_{ijk,i'j'k'} = & \sum_{l=1}^{p+1} \omega_l \hat{p}_i^0(x_l)\hat{p}_{i'}^0(x_l) \\
& \times \sum_{m=1}^{p+1} \omega_m \left(\frac{1-x_m}{2}\right)^{i+i'+1} \hat{p}_{j'}^{2i'-1}(x_m)\hat{p}_j^{2i-1}(x_m) \\
& \times \sum_{n=1}^{p+1} \omega_n \left(\frac{1-x_n}{2}\right)^{i+j+i'+j'+2} p_k^{2i+2j-2}(x_n)p_{k'}^{2i'+2j'-2}(x_n),
\end{aligned}
$$

i.e., for all $(i,j,k,i',j',k') \notin \mathfrak{S}^p(ijk,i'j'k')$, cf. (10), (9). This is done by the following algorithm.

Algorithm 3.1 *1. Compute*

$$
h_{i;i'}^{(1)} = \sum_{l=1}^{p+1} \omega_l \hat{p}_i^0(x_l)\hat{p}_{i'}^0(x_l)
$$

for all $i,i' \in \mathbb{N}$ satisfying $|i-i'| \le 2$ and $i,i' \le p$.

2. *Compute*

$$
h_{i,j;i',j'}^{(2)} = \sum_{m=1}^{p+1} \omega_m \left(\frac{1-x_m}{2}\right)^{i+i'+1} \hat{p}_j^{2i-1}(x_m)\hat{p}_{j'}^{2i'-1}(x_m)
$$

for all $i,j,i',j' \in \mathbb{N}$ satisfying $|i-i'| \le 2$, $|i+j-i'-j'| \le 4$, $i+j \le p$ and $i'+j' \le p$.

3. *Compute*

$$
h_{\beta,k;\beta,'k'}^{(3)} = \sum_{n=1}^{p+1} \omega_n \left(\frac{1-x_n}{2}\right)^{\beta+\beta'+2} p_k^{2\beta-2}(x_n)p_{k'}^{2\beta'-2}(x_n)
$$

for all $k,k',\beta,\beta' \in \mathbb{N}$ satisfying $|\beta-\beta'| \le 4$, $|\beta+k-\beta'-k'| \le 4$, $\beta+k \le p$ and $\beta'+k' \le p$.

4. *For all $(i,j,k,i',j',k') \notin \mathfrak{S}^p(ijk,i'j'k')$, set*

$$
m_{ijk,i'j'k'} = h_{i;i'}^{(1)} h_{i,j;i',j'}^{(2)} h_{i+j,k;i'+j',k'}^{(3)}.
$$

The algorithm requires the numerical evaluation of Jacobi and integrated Jacobi polynomials at the Gaussian points x_l, $l = 1,\dots,p+1$. In the next subsection, we present an algorithm which computes the required values $\hat{p}_k^\alpha(x_l)$, $m = 1,\dots,p+1$, $k = 1,\dots,p$, $\alpha = 1,\dots,2p$ in $\mathcal{O}(p^3)$ operations.

4 Fast Evaluation of Integrated Jacobi Polynomials

The integrated Jacobi polynomials needed in the computation of $m_{ijk,i'j'k'}$ (7) are $\hat{p}_i^0(x), \hat{p}_j^{2i-1}(x)$ (progressing in odd steps with respect to the parameter α) and

$\hat{p}_k^{2i+2j-2}(x)$ (progressing in even steps with respect to the parameter α). For $i+j+k \le p$ with $i \ge 2$ and $j,k \ge 1$ this means that

$$\left[\hat{p}_i^0(x)\right]_{2\le i\le p}, [\hat{p}_j^3(x)]_{1\le j\le p}, \ldots, [\hat{p}_j^{2p-3}(x)]_{1\le j\le p},$$
$$[\hat{p}_k^4(x)]_{1\le k\le p}, \ldots, [\hat{p}_k^{2p-4}(x)]_{1\le k\le p}$$

are needed. Since one group proceeds in even, the other one in odd steps, the total of integrated Jacobi polynomials that are needed is

$$\hat{p}_n^a(x), \qquad 1 \le n \le p-3, \quad 3 \le a \le 2p-3,$$

if we consider the integrated Legendre polynomials separately. However, integrating both sides of (3) yields

$$\hat{p}_{n+1}^{\alpha-1}(x) = \frac{1}{2n+\alpha}\left((n+\alpha)\hat{p}_{n+1}^{\alpha}(x) - n\hat{p}_n^{\alpha}(x)\right),$$

valid for all $n \ge 0$. Using this relation starting from the integrated Jacobi polynomials of highest degree, i.e., $\alpha = 2i-1 = 2p-3$, the remaining Jacobi polynomials can be computed using only two elements of the previous row. Note that for the initial values $n = 1$ we have $\hat{p}_1^{\alpha}(x) = 1+x$ for all α. For assembling the polynomials of highest degree the three term recurrence (4) is used. Summarizing, the evaluation of the functions at the Gaussian points can be done in $\mathcal{O}(p^3)$ operations. This is optimal in the three-dimensional case, but not in the two-dimensional case.

5 Complexity of the Algorithm

The cost of the last three steps is $\mathcal{O}(p^3)$, the first step requires $\mathcal{O}(p^2)$ operations. Together with the evaluation of the Jacobi polynomials, the algorithm requires in total $\mathcal{O}(p^3)$ flops.

This algorithm uses only the sparsity structure (10). Since all matrices $\hat{\mathscr{I}}^{(m)}$, $m = 1,\ldots,21$, have a similar sparsity structure of the form (10), this algorithm can be extended to all ingredients which are required for assembling/computing the element stiffness matrix (6) for the Laplacian, see [3]. The algorithm can also be extended to mass matrices or matrices arising from the discretization of elliptic problems in $H(\mathrm{curl})$ and $H(\mathrm{div})$, see [4]. For two-dimensional problems, the third step of the algorithm is not necessary. However, the values $h^{(2)}_{i,j;i',j'}$ have to be computed. Since this requires $\mathcal{O}(p^3)$ floating point operations, the total cost in 2D is also $\mathcal{O}(p^3)$.

Acknowledgments The work has been supported by the FWF projects P20121, P20162, and P23484.

Bibliography

[1] Milton Abramowitz and Irene A. Stegun. *Handbook of mathematical functions with formulas, graphs, and mathematical tables*, volume 55 of *National Bureau of Standards Applied Mathematics Series*. For sale by the Superintendent of Documents, U.S. Government Printing Office, Washington, D.C., 1964.

[2] I. Babuska, M. Griebel, and J. Pitkäranta. The problem of selecting the shape functions for a p-type finite element. *Internat. J. Numer. Methods Engrg.*, 28(8):1891–1908, 1989.

[3] S. Beuchler and V. Pillwein. Sparse shape functions for tetrahedral p-FEM using integrated Jacobi polynomials. *Computing*, 80(4):345–375, 2007.

[4] S. Beuchler, V. Pillwein, and S. Zaglmayr. Sparsity optimized high order finite element functions for $h(div)$ on simplices. Technical Report 2010-07, RICAM, 2010. submitted.

[5] S. Beuchler and J. Schöberl. New shape functions for triangular p-FEM using integrated Jacobi polynomials. *Numer. Math.*, 103(3):339–366, 2006.

[6] Leszek Demkowicz, Jason Kurtz, David Pardo, Maciej Paszyński, Waldemar Rachowicz, and Adam Zdunek. *Computing with hp-adaptive finite elements. Vol. 2*. Chapman & Hall/CRC Applied Mathematics and Nonlinear Science Series. Chapman & Hall/CRC, Boca Raton, FL, 2008. Frontiers: three dimensional elliptic and Maxwell problems with applications.

[7] George Em Karniadakis and Spencer J. Sherwin. *Spectral/hp element methods for computational fluid dynamics*. Numerical Mathematics and Scientific Computation. Oxford University Press, New York, second edition, 2005.

[8] J.M. Melenk, K. Gerdes, and C. Schwab. Fully discrete hp-finite elements: Fast quadrature. *Comp. Meth. Appl. Mech. Eng.*, 190:4339–4364, 1999.

[9] Ch. Schwab. *p- and hp-finite element methods*. Numerical Mathematics and Scientific Computation. The Clarendon Press Oxford University Press, New York, 1998. Theory and applications in solid and fluid mechanics.

A Non-overlapping Quasi-optimal Optimized Schwarz Domain Decomposition Algorithm for the Helmholtz Equation

Y. Boubendir[1], X. Antoine[2], and C. Geuzaine[3]

[1] Department of Mathematical Sciences and Center for Applied Mathematics and Statistics, NJIT, Univ. Heights. 323 Dr. M. L. King Jr. Blvd, Newark, NJ 07102, USA. boubendi@njit.edu

[2] Institut Elie Cartan Nancy (IECN), Nancy University, INRIA Corida Team, B.P. 239, F-54506 Vandoeuvre-lès-Nancy Cedex, France. Xavier.Antoine@iecn.u-nancy.fr

[3] University of Liège, Department of Electrical Engineering and Computer Science, Montefiore Institute B28, B-4000 Liège, Belgium cgeuzaine@ulg.ac.be

1 Introduction

In this paper, we present a new non-overlapping domain decomposition algorithm for the Helmholtz equation. We are particularly interested in the method introduced by P.-L. Lions [6] for the Laplace equation and extended to the Helmholtz equation by B. Després [3]. However, this latest approach provides slow convergence of the iterative method due to the choice of the transmission conditions. Thus, in order to improve the convergence, several methods were developed [4, 5, 9, 10]. The main idea in [5, 9] consists in computing a more accurate approximation of the Dirichlet-to-Neuman (DtN) operator than the one proposed in [3] by using particular local transmission conditions. We propose in this work a different approach to approximate the DtN map. We mainly use Padé approximants to suitably localize the nonlocal representation of the DtN operator [8, 11]. This results in an algorithm with quasi-optimal convergence properties.

2 Model Problem and Non-overlapping Domain Decomposition Method

For the sake of simplicity, we limit ourselves to the evaluation of the two-dimensional time-harmonic scattering wave by an obstacle denoted by K. The three-dimensional case is treated similarly without adding any difficulty. We consider the model problem given by the system

R. Bank et al. (eds.), *Domain Decomposition Methods in Science and Engineering XX*, Lecture Notes in Computational Science and Engineering 91,
DOI 10.1007/978-3-642-35275-1_61,

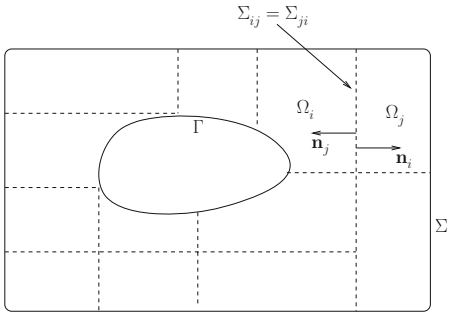


Fig. 1. Example of 2D non-overlapping domain decomposition method

$$\begin{cases} \Delta u + k^2 u = 0 \quad \text{in } \mathbb{R}^2 \backslash K, \\ u = f \quad \text{on } \Gamma = \partial K, \\ \lim\limits_{|x| \to \infty} |x|^{1/2} (\partial_{|x|} u - iku) = 0, \end{cases} \tag{1}$$

composed of the Helmholtz equation, the Dirichlet condition on Γ (TE polarization in electromagnetics) where $f = -e^{ik\alpha \cdot x}$ describes the incident plane wave with $|\alpha| = 1$ and k is the wavenumber, and the Sommerfeld radiation condition. To solve (1), we combine the absorbing boundary condition method [1, 2] with non-overlapping domain decomposition methods. The absorbing boundary conditions method consists of truncating the computational domain using an artificial interface Σ, and reducing the system (1) to the following one

$$\begin{cases} \Delta u + k^2 u = 0 \quad \text{in } \Omega, \\ u = f \quad \text{on } \Gamma, \\ \partial_{\mathbf{n}} u + \mathscr{B} u = 0 \quad \text{on } \Sigma, \end{cases} \tag{2}$$

where Ω is the bounded domain enclosed by Σ and Γ, $\mathscr{B}$ indicates the approximation of the Dirichlet-to-Neuman (DtN) operator, and $\mathbf{n}$ is the outward normal to Σ. We are interested in the domain decomposition method introduced in [3, 6]. The first step of this approach consists in splitting Ω into several subdomains Ω_i, $i = 1, \ldots, N$, such that

- $\overline{\Omega} = \bigcup_{i=1}^{N} \overline{\Omega}_i$ $(i = 1, \ldots, N)$,
- $\Omega_i \cap \Omega_j = \emptyset$, if $i \neq j$, $(i, j = 1, \ldots, N)$,
- $\partial\Omega_i \cap \partial\Omega_j = \overline{\Sigma}_{ij} = \overline{\Sigma}_{ji}$ $(i, j = 1, \ldots, N)$ is the artificial interface (see Fig. 1) separating Ω_i from Ω_j as long as its interior Σ_{ij} is not empty.

Then, applying the Lions-Després algorithm, the solution of the initial problem (1) is reduced to an iterative procedure, where each iteration is performed by solving the local problems

$$\begin{cases} \Delta u_i^{(n+1)} + k^2 u_i^{(n+1)} = 0 \quad \text{in } \Omega_i, \\ u_i^{(n+1)} = f_i \quad \text{on } \Gamma_i, \\ \partial_{\mathbf{n}_i} u_i^{(n+1)} + \mathscr{B} u_i^{(n+1)} = 0 \quad \text{on } \Sigma_i \end{cases} \tag{3a}$$

$$\partial_{\mathbf{n}_i} u_i^{(n+1)} + \mathscr{S} u_i^{(n+1)} = g_{ij}^{(n)} \quad \text{on } \Sigma_{ij}, \tag{3b}$$

and forming the quantities to be transmitted through the interfaces

$$g_{ij}^{(n+1)} = -\partial_{\mathbf{n}_j} u_j^{(n+1)} + \mathscr{S} u_j^{(n+1)} = -g_{ij}^{(n)} + 2\mathscr{S} u_j^{(n+1)} \quad \text{on } \Sigma_{ij}, \tag{4}$$

where $u_i = u|_{\Omega_i}$, $\mathbf{n}_i$ (resp. $\mathbf{n}_j$) is the outward unit normal of the boundary of Ω_i (resp. Ω_j), $i = 1,\ldots,N$, $j = 1,\ldots,N$, $\Gamma_i = \partial\Omega_i \cap \Gamma$ and $\Sigma_i = \partial\Omega_i \cap \Sigma$. Note that the boundary condition on Γ_i (resp. Σ_i) does not take place if the interior of $\partial\Omega_i \cap \Gamma$ (resp. $\partial\Omega_i \cap \Sigma$) is the empty set.

3 New Transmission Conditions

It is well established that the convergence of the domain decomposition algorithms depends on the choice of the transmission operator $\mathscr{S}$. In the original method proposed by B. Després [3], the usual approximation of the DtN operator $\mathscr{S} u = -\imath k u$ is used. The resulting algorithm does not treat efficiently the evanescent modes of the iteration operator which impairs the iterative method [9]. In order to improve the convergence, two techniques, based on the modification of the operator $\mathscr{S}$, were proposed. First, the optimized Schwarz method introduced by Gander et al. [5]. It consists of using local second-order approximations of the DtN operator $\mathscr{S} u = \delta u + \gamma \partial_s^2 u$, where ∂_s is the tangential derivative operator, and the coefficients δ and γ are optimized using the rate of convergence obtained in the case of the half-plane. The second method, called the "evanescent modes damping algorithm" (EMDA), was introduced by Boubendir et al. [9, 10]. In this case, S is chosen as $\mathscr{S} u = -\imath k u + \mathscr{X} u$ where $\mathscr{X}$ is a self-adjoint positive operator. We only consider here the usual case where $\mathscr{X}$ is a real-valued positive coefficient. In this paper we propose a new "square-root" transmission operator [7, 8, 11] that takes the following form:

$$\mathscr{S} u = -\imath k \mathrm{Op}\left(\sqrt{1 - \frac{\xi^2}{k_\varepsilon^2}}\right) u, \tag{5}$$

where

$$k_\varepsilon = k + \imath\varepsilon \tag{6}$$

is a complexified wavenumber, and the notation $\sqrt{z}$ designates the principal determination of the square-root of a complex number z with branch-cut along the negative real axis. This choice of the square-root operator is motivated by developments of absorbing boundary conditions (ABC) for scattering problems [1, 2]. Generally speaking, the usual techniques to develop absorbing boundary conditions consists

mainly in using Taylor expansions to approximate the symbol of the DtN operator. However, these approximations prevent the modelling of the three parts describing the wave (propagating, evanescent and transition) at the same time, which affects, in return, the final accuracy of the solution. This problem can be solved by high-order local ABC introduced in [7, 8], which uses (5) to model all the scattering modes: propagating, evanescent as well as (in an approximate way) grazing. The localization is performed with complex Padé approximants, and the coefficient ε in (6) can then be chosen to minimize spurious reflections at the boundary. In the context of domain decomposition methods, this optimization of ε improves the spectrum of the iteration operator on these grazing modes. As it is shown in [8], the optimal value of this parameter is given by $\varepsilon = 0.4k^{1/3}\mathscr{H}^{2/3}$, where $\mathscr{H}$ is the mean curvature on the interface.

4 Localization of the Square-Root Operator Using Padé Approximants

Because the square-root operator (5) is nonlocal, its use in the context of finite element method is ineffective since it would lead to consider full matrices for the transmission boundaries. A localization process of this operator can be efficiently done by using partial differential (local) operators and obtain sparse matrices. This is performed [7, 8, 11] in rotating branch-cut approximation of the square-root and then applying complex Padé approximants of order N_p,

$$\begin{aligned}\sqrt{1-\frac{\xi^2}{k_\varepsilon^2}}u &\approx R^{\alpha}_{N_p}(-\frac{\xi^2}{k_\varepsilon^2})u \\ &= C_0 u + \sum_{\ell=1}^{N_p} A_\ell(\frac{-\xi^2}{k_\varepsilon^2})(1+B_\ell(\frac{-\xi^2}{k_\varepsilon^2}))^{-1}u,\end{aligned} \tag{7}$$

which correspond to the complex Padé approximation

$$\sqrt{1+z} \approx R^{\alpha}_{N_p}(z) = C_0 + \sum_{\ell=1}^{N_p} \frac{A_\ell z}{1+B_\ell z}, \tag{8}$$

and where the complex coefficients C_0, A_ℓ and B_ℓ are given by

$$C_0 = e^{\imath\frac{\alpha}{2}} R_{N_p}(e^{-\imath\alpha}-1), A_\ell = \frac{e^{-\frac{\imath\alpha}{2}}a_\ell}{(1+b_\ell(e^{-\imath\alpha}-1))^2}, B_\ell = \frac{e^{-\imath\alpha}b_\ell}{1+b_\ell(e^{-\imath\alpha}-1)}.$$

Here, α is the angle of rotation, (a_ℓ, b_ℓ), $\ell = 1,\ldots,N_p$, are the standard real Padé coefficients

$$a_\ell = \frac{2}{2N_p+1}\sin^2(\frac{\ell\pi}{2N_p+1}),\ b_\ell = \cos^2(\frac{\ell\pi}{2N_p+1}), \tag{9}$$

and R_{N_p} is the real Padé approximant of order N_p

$$\sqrt{1+z} \approx R_{N_p}(z) = 1 + \sum_{\ell=1}^{N_p} \frac{a_\ell z}{1+b_\ell z}. \tag{10}$$

For a variational representation, the approximation of the Padé-localized square-root transmission operators is realized by using auxiliary coupled functions [7, 11]

$$\mathscr{S} u = -\imath k (C_0 u + \sum_{\ell=1}^{N_p} A_\ell \mathrm{div}_{\Sigma_d}(\frac{1}{k_\varepsilon^2} \nabla_{\Sigma_d} \varphi_\ell)) \quad \text{on } \Sigma_d, \tag{11}$$

where the functions φ_ℓ, $\ell = 1,..,N_p$, are defined on any artificial interface Σ_d as the solutions of the surface PDEs

$$(1 + B_\ell \mathrm{div}_{\Sigma_d}(\frac{1}{k_\varepsilon^2} \nabla_{\Sigma_d}))\varphi_\ell = u. \tag{12}$$

The resulting transmitting condition is a Generalized Impedance Boundary Condition, and is denoted by GIBC(N_p, α, ε) for the Padé approximation with N_p auxiliary functions, for an angle of rotation α and a damping parameter ε. The lowest-order approximation $\mathscr{S} = -\imath k I$ (resp. $\mathscr{S} = -\imath k u + \mathscr{X} u$) is denoted by IBC(0) (resp. IBC($\mathscr{X}$)).

5 Numerical Results

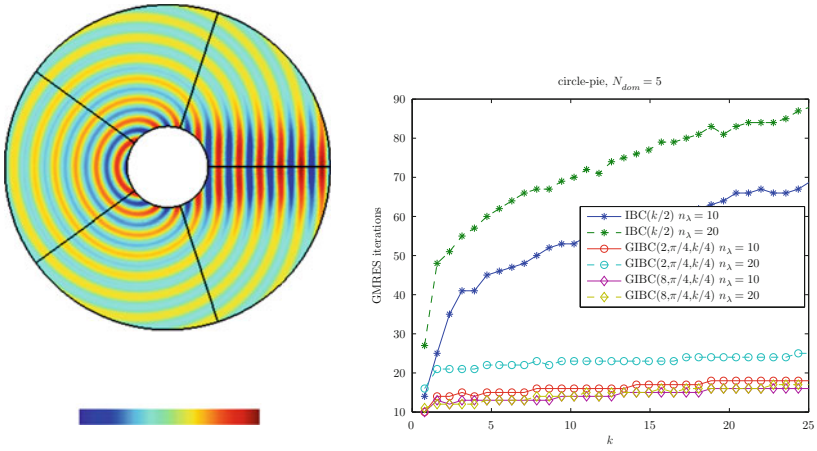


Fig. 2. *Left*: decomposition of the computational domain. *Right*: iteration number with respect to the wavenumber k for two densities of discretization n_λ

The numerical tests presented here concern the scattering of a plane wave by a unit sound-soft circular cylinder. We truncate the computational domain using a circle of radius equal to 4, on which the second-order Bayliss-Turkel absorbing condition [1] is set (see problem (2)). We perform these numerical tests on partitions of the type displayed in Fig. 2, and we refer to them as "circle-pie". We use a finite element method with linear (P1) basis functions to approximate the solution in each subdomain. The implementation of this method with Padé approximants is described in [11]. The iterative problem is solved using GMRES and the iterations are stopped when the initial residual has decreased by a factor of 10^{-6}.

We begin by testing the iterative method with respect to the wavenumber k. Let us consider the number of subdomains $N_{\text{dom}} = 5$. Because the interfaces are straight, as depicted on the left picture of Fig. 2, ε cannot be optimized as described in Sect. 3. However, numerical simulations show that $\varepsilon = k/4$ is an appropriate choice for this kind of interfaces. On the right picture of Fig. 2, we represent the behavior of the number of iterations. We choose two densities of discretization points per wavelength n_λ. We compare the new algorithm noted GIBC(N_p, $\pi/4$, ε), where N_p is the Padé number and $\pi/4$ the angle of rotation, with the EMDA algorithm designated by IBC($k/2$). In this latest case, the number of iterations clearly increases with respect to k and n_λ. However, for GIBC(N_p, $\pi/4$, ε), the convergence rate is almost independent of both the wavenumber and density of discretization points per wavelength. In particular, the convergence for $N_p = 2$ and $N_p = 8$ is similar. This means that the cost of the solution when solving local problems is comparable to the other methods with usual local transmission conditions (see [11] for more details).

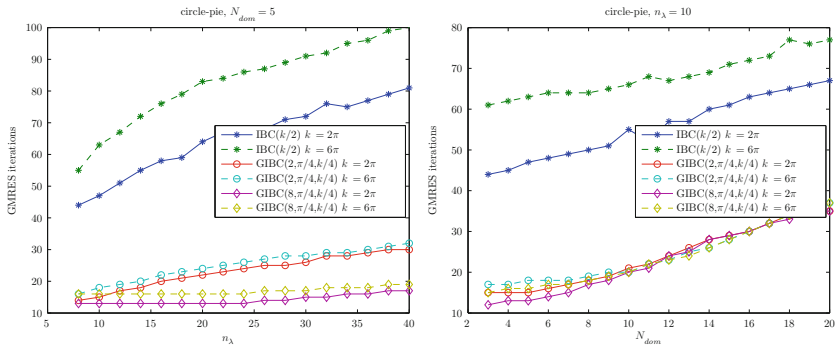


Fig. 3. Number of iterations with respect to the density of discretization n_λ and the number of subdomains N_{dom}

In Fig. 3, we show the number of iterations with respect to: (i) the density of discretization points per wavelength n_λ for two wavenumbers k, and (ii) the number of subdomains N_{dom}. We can see that for a small Padé number ($N_p = 2$), the convergence is almost independent of the mesh size. A larger choice of N_p will provide an optimal result. We also see that the number of iterations with respect to the number

of subdomains does not deteriorate with increasing values of N_p or k, contrary to IBC($k/2$).

6 Conclusion

We designed in this paper a new non-overlapping domain decomposition algorithm for the Helmholtz equation with quasi-optimal convergence properties. It is based on a suitable approach which consists in using Padé approximants to approximate the DtN operator. The analysis of this new approach can be found in [11], as well as several numerical tests including the three-dimensional case.

Acknowledgments Y. Boubendir gratefully acknowledges support from NSF through grant No. DMS–1016405. X. Antoine gratefully acknowledges support from the Agence Nationale pour la Recherche (Ref: ANR-09-BLAN-0057-01) and the Fondation de Recherche pour l'Aéronautique et l'Espace (IPPON Project). C. Geuzaine gratefully acknowledges support from the Belgian Science Policy (IAP P6/21), Belgian French Community (ARC 09/14-02) and Walloon Region (WIST3 No 1017086 "ONELAB").

Bibliography

[1] A. Bayliss, M. Gunzburger and E. Turkel. Boundary Conditions for the Numerical Solutions of Elliptic Equations in Exterior Regions. *SIAM Journal of Applied Mathematics*, 42:430–451, 1982.

[2] B. Engquist and A. Majda. Absorbing Boundary Conditions for the Numerical Simulation of Waves. *Mathematics of Computation*, 23:629–651, 1977.

[3] Bruno Després. Domain decomposition method and the Helmholtz problem.II. In *Second International Conference on Mathematical and Numerical Aspects of Wave Propagation (Newark, DE, 1993)*, pages 197–206, Philadelphia, PA, 1993. SIAM.

[4] C. Farhat, A. Macedo, M. Lesoinne, F.X. Roux, F. Magoulès and A. De La Bourdonnaye. Two-Level Domain Decomposition Methods with Lagrange Multipliers for the Fast Iterative Solution of Acoustic Scattering Problems. *Computer Methods in Applied Mechanics and Engineering*, 184:213–239, 2000.

[5] M. Gander, F. Magoulès, F. Nataf. Optimized Schwarz Methods without Overlap for the Helmholtz Equation. *SIAM J. Sci. Comput.*, 24(1):38–60, 2002.

[6] P.-L. Lions. On the Schwarz Alternating Method III: A Variant for Non Overlapping Subdomains. In *Third International Symposium on Domain Decomposition Methods for Partial Differential Equations*, pages 20–22, Philadelphia, PA, 1990. SIAM.

[7] R. Kerchroud, X. Antoine and A. Soulaimani. Numerical Accuracy of a Padé-Type Non-Reflecting Boundary Condition for the Finite Element Solution of Acoustic Scattering Problems at High-Frequency. *International Journal for Numerical Methods in Engineering*, 64:1275–1302, 2005.

[8] X. Antoine, M. Darbas, and Y.Y. Lu. An Improved Surface Radiation Condition for High-Frequency Acoustics Scattering Problems. *Computer Methods in Applied Mechanics and Engineering*, 195:4060–4074, 2006.

[9] Y. Boubendir. An Analysis of the BEM-FEM Non-Overlapping Domain Decomposition Method for a Scattering Problem. *Journal of Computational and Applied Mathematics*, 204(2):282–291, 2007.

[10] Y. Boubendir, A. Bendali and M. B. Fares. Coupling of a Non-Overlapping Domain Decomposition Method for a Nodal Finite Element Method with a Boundary Element Method. *International Journal for Numerical Methods in Engineering*, 73(11):1624–1650, 2008.

[11] Y. Boubendir, X. Antoine and C. Geuzaine. A Quasi-Optimal Non-Overlapping Domain Decomposition Algorithm for the Helmholtz Equation. *Journal of Computational Physics*, 231:262–280, 2012.

A Continuous Approach to FETI-DP Mortar Methods: Application to Dirichlet and Stokes Problem

E. Chacón Vera[1], D. Franco Coronil[1] and A. Martínez Gavara[2]

[1] Dpto. Ecuaciones Diferenciales y Análisis Numérico, Facultad de Matemáticas, Universidad de Sevilla, Tarfia sn. 41012 Sevilla, SPAIN, email: {eliseo,franco}@us.es
[2] Dpto. de Estadística e Investigación Operativa, Universidad de Valencia, Valencia, SPAIN, email: Ana.Martinez-Gavara@uv.es

Summary. In this contribution we extend the FETI-DP mortar method for elliptic problems introduced by Bernardi et al. [2] and Chacón Vera [3] to the case of the incompressible Stokes equations showing that the same results hold in the two dimensional setting. These ideas extend easily to three dimensional problems. Finally some numerical tests are shown as a conclusion. This contribution is a condensed version of a more detailed forthcoming paper. We use standard notation, see for instance [1].

1 Incompressible Stokes Equations

Let $\Omega \subset \mathbb{R}^2$ be a polygonal domain. We look for $u \in \mathbf{H}_0^1(\Omega) = (H_0^1(\Omega))^2$ and $p \in L^2(\Omega)$ such that $\int_\Omega p = 0$ and

$$\begin{aligned}(\nabla u, \nabla v)_\Omega - (p, div(v))_\Omega &= (f,v)_\Omega, \quad \forall v \in \mathbf{H}_0^1(\Omega)\\ -(q, div(u))_\Omega &= 0, \qquad \forall q \in L^2(\Omega).\end{aligned}$$

We better accomodate the restriction on the pressure by adding a new scalar unknown: we look for a pair of values $(u,\tau) \in \mathbf{H}_0^1(\Omega) \times \mathbb{R}$ and $p \in L^2(\Omega)$ such that

$$\begin{aligned}(\nabla u, \nabla v)_\Omega - (p, div(v))_\Omega + t\,(\tau - \int_\Omega p) &= (f,v)_\Omega, \quad \forall (v,t) \in \mathbf{H}_0^1(\Omega) \times \mathbb{R}\\ -(q, div(u))_\Omega - \tau \int_\Omega q &= 0, \qquad \forall q \in L^2(\Omega).\end{aligned}$$

Set $W = \mathbf{H}_0^1(\Omega) \times \mathbb{R}$ normed by $\|\underline{v}\|_W^2 = \|(v,t)\|_W^2 = \|\nabla v\|_{0,\Omega}^2 + t^2$ for any $\underline{v} = (v,t) \in W$, let $(\cdot,\cdot)_W$ be the scalar product on W and $b : W \times L^2(\Omega) \mapsto \mathbb{R}$ given by

$$b(q,(v,t)) = -(q, div(v))_\Omega - t \int_\Omega q.$$

Then, we look for $\underline{u} = (u,\tau) \in W$ and $p \in L^2(\Omega)$ such that

R. Bank et al. (eds.), *Domain Decomposition Methods in Science and Engineering XX*, Lecture Notes in Computational Science and Engineering 91,
DOI 10.1007/978-3-642-35275-1_62,

$$(\underline{u},\underline{v})_W + b(p,\underline{v}) = (f,v)_\Omega, \quad \forall \underline{v} \in W \tag{1}$$

$$b(q,\underline{u}) = 0, \quad \forall q \in L^2(\Omega). \tag{2}$$

It is quite straightforward to see that:

Lemma 1. *There exists a positive constant $\beta > 0$ such that for all $p \in L^2(\Omega)$*

$$\sup_{(v,t)\in W} \frac{b(p,(v,t))}{\|(v,t)\|_W} \geq \sup_{v\in \mathbf{H}_0^1(\Omega), t\in\mathbb{R}} \frac{b(p,(v,t))}{(\|\nabla v\|_{0,\Omega}^2 + t^2)^{1/2}} \geq \beta \|p\|_{0,\Omega}. \tag{3}$$

As a consequence, problem (1)–(2) is well posed and its unique solution is the one of the original Stokes problem with Dirichlet homogeneous boundary conditions.

Next, we split $\Omega = \cup_{s=1}^S \Omega^s$ with nonoverlaping polygonal subdomains, suppose that

$$\Gamma_{s,t} = \partial\Omega^s \cap \partial\Omega^t$$

is either an edge (i.e., a segment), a crosspoint or empty and, finally, consider $\mathscr{E}_0 = \{\Gamma_e\}_{e=1,..,E}$ the sorted set of all edges inside Ω. We suppose that each Ω^s is of area $\mathscr{O}(H^2)$ and shape regular while each Γ_e is of length $\mathscr{O}(H)$ for some fixed $H > 0$. The set of all vertices of the polygonal subdomains Ω^s that are not on $\partial\Omega$ will be called **cross points** and denoted by $\mathscr{C}$. Finally, we denote by $[v]_{\Gamma_e}$ the jump across any interface Γ_e.

We take

$$X_\delta = \{v \in L^2(\Omega); v^s = v_{|_{\Omega^s}} \in H^1(\Omega^s) \cap H_0^1(\Omega),\ 1 \leq s \leq S\},$$

$$X = \{v \in X_\delta,\ [v]_{\Gamma_e} \in H_{00}^{1/2}(\Gamma_e),\ \forall \Gamma_e \in \mathscr{E}_0\}.$$

With $\mathbf{X} = X \times X$ we construct $\mathbf{V} = \mathbf{X} \times \mathbb{R}$ and represent by $\underline{v} = (v,t)$ any element of $\mathbf{V}$ where $v \in \mathbf{X}$ and $t \in \mathbb{R}$. $\mathbf{V}$ is Hilbert space with norm $\|\underline{v}\|_{\mathbf{V}}^2 = |v|_{\mathbf{X}}^2 + t^2$ where, thanks to Poincaré's inequality, the norm of v is

$$|v|_{\mathbf{X}} = \{\sum_{s=1}^S \|\nabla v^s\|_{0,\Omega^s}^2 + \sum_{e=1}^E \|[v]_{\Gamma_e}\|_{1/2,00,\Gamma_e}^2\}^{1/2}.$$

Here, $\|\cdot\|_{1/2,00,\Gamma_e}$ is the norm induced by the scalar product $(\cdot,\cdot)_{1/2,00,\Gamma_e}$ on $H_{00}^{1/2}(\Gamma_e)$, see [5]. To simplify, let $\{\cdot,\cdot\}_{\Gamma_e} = (\cdot,\cdot)_{1/2,00,\Gamma_e}$. For the pressure space we consider $\mathbf{M} = \prod_{s=1}^S L^2(\Omega^s) (\approx L^2(\Omega))$ and define the continuous bilinear form $b : \mathbf{M} \times \mathbf{V} \mapsto \mathbb{R}$ given by

$$b(q,\underline{v}) = -\sum_{s=1}^S (q^s, div(v^s))_{\Omega^s} - t\sum_{s=1}^S \int_{\Omega^s} q^s, \quad \forall q^s \in L^2(\Omega^s).$$

Next, for each $\Gamma_e \in \mathscr{E}_0$ we take $\mathbf{H}_{00}^{1/2}(\Gamma_e) = (H_{00}^{1/2}(\Gamma_e))^2$, and handle the Lagrange multipliers for the jumps with the space $\mathbf{N} = \prod_{e=1}^E \mathbf{H}_{00}^{1/2}(\Gamma_e)$.

We propose to look for $\underline{u} = (u, \tau) \in \mathbf{V}$, $p = \{p^s\}_s \in \mathbf{M}$ and $\lambda = \{\lambda_e\}_e \in \mathbf{N}$ such that

$$\begin{aligned}
&\sum_{s=1}^{S} (\nabla u^s, \nabla v^s)_{\Omega^s} + \sum_{e=1}^{E} \{[u]_{\Gamma_e}, [v]_{\Gamma_e}\}_{\Gamma_e} + \tau t \\
&- \sum_{s=1}^{S} (p^s, div(v^s))_{\Omega^s} - t \sum_{s=1}^{S} \int_{\Omega^s} p^s + \sum_{e=1}^{E} \{\lambda_e, [v]_{\Gamma_e}\}_{\Gamma_e} = \sum_{s=1}^{S} (f, v^s)_{\Omega^s}, \\
&- \sum_{s=1}^{S} (q^s, div(u^s))_{\Omega^s} - \tau \sum_{s=1}^{S} \int_{\Omega^s} q^s = 0, \\
&\sum_{e=1}^{E} \{\mu_e, [u]_{\Gamma_e}\}_{\Gamma_e} = 0
\end{aligned}$$

for all $\underline{v} = (v, t) \in \mathbf{V}$, $q = \{q^s\}_s \in \mathbf{M}$ and $\mu = \{\mu_e\}_e \in \mathbf{N}$.

We see that we added the jumps to the elliptic terms and replaced the pairings $H_{00}^{-1/2}(\Gamma) - H_{00}^{1/2}(\Gamma)$ for the normal fluxes on the edges by the scalar product in $H_{00}^{1/2}(\Gamma)$. As a consequence, we have made a regularization of order 1 for the Lagrange multipliers and now all terms are suitable to compute in a Galerkin approach. Moreover, the solution to this problem is that of the incompressible Stokes equations on Ω.

Next, we elliminate via a standard Schur process the primal variables $\underline{u}$ and p in terms of the dual variable λ, and obtain a dual problem that once solved will give the correct boundary data for the primal variables. Thanks to the fact that the elliptic part is the scalar product on $\mathbf{V}$, that the inf-sup condition for the bilinear form b is achieved with velocities without jumps and that the inf-sup condition for c is achieved with velocities with jumps, our dual problem is a well posed symmetric positive definite problem.

2 Finite Dimensional Approach

We consider a conforming triangulation $\mathscr{T}_h$, h is the mesh size, of $\overline{\Omega}$ that contains the skeleton $\mathscr{E}_0$ as union of edges of triangles and such that on each edge only one partition is inherited from both sides. As $\mathscr{T}_h$ is also compatible with the subdivision of Ω, its restriction to each $\overline{\Omega_s}$ gives a mesh $\mathscr{T}_h^s$ on $\overline{\Omega^s}$. We use the Taylor-Hood finite element for the velocity and pressure pair on each subdomain. Define the family of subspaces $\{Y_h\}_h \subset H_0^1(\Omega)$ and $\{Q_h\}_h \subset H^1(\Omega)$ given by

$$\begin{aligned}
Y_h &= \{v \in H_0^1(\Omega);\ v_{|_\kappa} \in \mathbb{P}_2(\kappa),\ \forall \kappa \in \mathscr{T}_h\}, \\
Q_h &= \{p \in H^1(\Omega);\ p_{|_\kappa} \in \mathbb{P}_1(\kappa),\ \forall \kappa \in \mathscr{T}_h\}
\end{aligned}$$

where $\mathbb{P}_r(\kappa)$ is the space of polynomials of degree less or equal to r in the two variables x and y. On each subdomain, we take also

$$Y_h(\Omega^s) = Y_h \cap H^1(\Omega^s), \quad Q_h(\Omega^s) = Q_h \cap H^1(\Omega^s),\ s \le S.$$

Consider now $\mathbf{X}_h = X_h \times X_h$, where X_h is the broken version of Y_h given by

$$X_h = \{v \in L^2(\Omega);\ v^s \in Y_h^s,\ \forall\, s = 1,2,\ldots,S, \\ \text{and } v \text{ is continuous at every cross point in } \mathscr{C}\} \subset X,$$

define $\mathbf{V}_h = \mathbf{X}_h \times \mathbb{R}$, $\mathbf{M}_h = \prod_{s=1}^{S} Q_h(\Omega^s)$ and finally $\mathbf{N}_h \subset \mathbf{N}$ is given by the restriction of functions in $\mathbf{X}_h$ to the skeleton $\mathscr{E}_0$.

The discrete uniform inf-sup condition for c on the pair $\mathbf{V}_h$ and $\mathbf{N}_h$ is by now a well known result and the discrete uniform inf-sup condition for b is a consequence of Theorem 1.12 pp. 130 in [4]. The idea is to use locally on each subdomain Ω^s the stability of the pair $\mathbb{P}_2 - \mathbb{P}_1$ and that of the pair $\mathbb{P}_2 - \mathbb{P}_0$ globally on the substructures Ω^s of Ω. This inf-sup condition is achieved with a discrete continuous function in the wohle of Ω and, as a consequence, the continuous setting is replicated and the equation for the multiplier can be solved via Conjugate Gradient Method (CG) without preconditioner. Then, we have

1. An external computational cicle, the CG for the Lagrange multiplier with a fixed number of iterations independent of the discretization parameter h and
2. At each iteration of this external cicle, the resolution of a primal problem of the form:
   Find $(\underline{w}_h, q_h) \in \mathbf{V}_h \times \mathbf{M}_h$ such that

$$(\underline{w}_h, \underline{v}_h)\mathbf{v} + b(q_h, \underline{v}_h) = (\xi, \underline{v}_h) \quad \forall \underline{v}_h \in \mathbf{V}_h, \\ b(p, \underline{w}_h) = 0 \quad \forall p \in \mathbf{M}_h$$

   where for the initial residuous r_0 we have $(\xi, \underline{v}_h) = \sum_{s=1}^{S} (f, v_h^s)_{\Omega^s}$ and for the iteration $m \ge 0$ we have $(\xi, \underline{v}_h) = \sum_{e=1}^{E} \{\{d_m\}_e, [v_h]_{\Gamma_e}\}_{\Gamma_e} = 0$

A closer inspection to the general form of this saddle point problem for the primal variables shows that the solution can be obtained by means of independent solves per subdomain. Ordering the unknows per subdomains, $x^s = (u^s, p^s)$ and $x^C = u^C$, the linear system for the primal variables is

$$\begin{pmatrix} M_{11} & M_{1,2} & \ldots & \ldots & \ldots & M_{1,S} & M_{1,C} & D_1 \\ M_{21} & M_{2,2} & M_{2,3} & \ldots & \ldots & \ldots & M_{2,C} & D_2 \\ M_{31} & M_{3,2} & M_{3,3} & M_{3,4} & \ldots & \ldots & M_{3,C} & D_3 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots & \vdots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots & \vdots & \vdots \\ \vdots & \ldots & \ldots & \ldots & M_{S,S-1} & M_{S,S} & M_{S,C} & D_S \\ M_{1,C}^t & M_{2,C}^t & \ldots & \ldots & M_{S-1,C}^t & M_{S,C}^t & M_{C,C} & 0 \\ D_1^t & D_2^t & \ldots & \ldots & D_{S-1}^t & D_S^t & 0^t & 1 \end{pmatrix} \begin{pmatrix} x^1 \\ x^2 \\ x^3 \\ \vdots \\ \vdots \\ x^S \\ x^C \\ \tau \end{pmatrix} = \begin{pmatrix} b^1 \\ b^2 \\ b^3 \\ \vdots \\ \vdots \\ b^S \\ b^C \\ 0 \end{pmatrix}$$

where the different blocks are of the form

$$M_{s,s} = \begin{pmatrix} A_{s,s} & B_{s,s} \\ B_{s,s}^t & 0 \end{pmatrix}, \; M_{s,s'} = \begin{pmatrix} A_{s,s'} & 0 \\ 0 & 0 \end{pmatrix}, \; M_{s,C} = \begin{pmatrix} A_{s,C} \\ B_{s,C}^t \end{pmatrix}, \; M_{C,C} = A_{C,C}$$

here each block $M_{s,s}$ is similar to a standard Stokes matrix on the subdomain Ω^s, but with our interface contributions, each block $M_{s,s'}$ is sparse and contains the interaction through interfaces of the domain Ω^s with $\Omega^{s'}$, the rectangular blocks $M_{s,C}$ contains the interaction with the crosspoints and $M_{C,C}$ contains the interaction of the crosspoints with themselves. Although this linear system couples all the subdomains it can be solved by means of the Preconditioned Conjugate Gradient Method using as a preconditioner the matrix P formed by the main blocks

$$P = \begin{pmatrix} M_{11} & 0 & \dots & \dots & 0 & M_{1,C} & D_1 \\ 0 & M_{2,2} & 0 & \dots & 0 & M_{2,C} & D_2 \\ 0 & 0 & M_{3,3} & 0 & \ddots & M_{3,C} & D_3 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots & \vdots \\ \dots & \dots & \dots & 0 & M_{S,S} & M_{S,C} & D_S \\ M_{1,C}^t & M_{2,C}^t & \dots & M_{S-1,C}^t & M_{S,C}^t & M_{C,C} & 0 \\ D_1^t & D_2^t & \dots & D_{S-1}^t & D_S^t & 0^t & 1 \end{pmatrix}.$$

Therefore, the main task here is the resolution of a linear system of the form $Px = b$ which is done using a Schur complement process for the variables x^C and τ. The equations are

$$(M_{C,C} - \sum_{s=1}^{S} M_{s,C}^t M_{s,s}^{-1} M_{s,C}) x^C - \sum_{s=1}^{S} M_{s,C}^t M_{s,s}^{-1} D_s \tau = b^C - \sum_{s=1}^{S} M_{s,C}^t M_{s,s}^{-1} b^s,$$
$$\sum_{s=1}^{S} D_s^t M_{s,s}^{-1} M_{s,C} x^C + (\sum_{s=1}^{S} D_s^t M_{s,s}^{-1} D_s - 1)\tau = \sum_{s=1}^{S} D_s^t M_{s,s}^{-1} b^s.$$

We finally write x^C in terms of τ and solve first for τ, next x^C and finally compute all the x^s. As a consequence, the main job is performed with independent solves of the matrices $M_{s,s}$ that can be performed independently, i.e., computations of the form

$$M_{s,s}^{-1} b^s, \qquad M_{s,s}^{-1} M_{s,C}, \qquad M_{s,s}^{-1} D_s.$$

3 Some Numerical Tests

For $L = 1, 2, 3, \dots$ integer we consider on $\Omega_L = [0, L] \times [0, 1]$ the exact solution

$$u(x,y) = \begin{pmatrix} -\sin^3(\pi x L^{-1}) \sin^2(\pi y) \cos(\pi y) \\ -L^{-1} \sin^2(\pi x L^{-1}) \sin^3(\pi y) \cos(\pi x L^{-1}) \end{pmatrix}, \quad p(x,y) = \frac{x^2}{L^2} - y^2$$

and partition Ω_L into $\Omega_L^s = (s-1, s) \times (0, 1)$ for $s = 1, 2, \dots, L$. For the dual problem we start our iteration process with $\lambda_{0,e} = 0$ on each Γ_e and stop all iterations according

to a relative residual less than 10^{-6}. In this example the gradients control the jumps and there is no need to introduce them in the elliptic part; then the blocks $M_{s,t}$ are null for $s \neq t$. Then, there is no need for a PCG in the internal cycle. The following Table 1 shows that the iteration count for the dual problem is mesh independent on different configurations Table 2 shows relative errors with respect to the true solution

	h = 1/24	h = 1/48	h = 1/96
L = 4	17	17	17
L = 8	23	24	24
L = 16	37	39	39

Table 1. Mesh independent iteration count for the dual problem on different configurations and for different values of h on $\Omega_L = [0,L] \times [0,1]$. The number of subdomains is L given by $\Omega^s = [s-1,s] \times [0,1]$ for $s = 1,2,3,\ldots,L$

u and p on Ω_L Finally, we take on $\Omega = (0,1)^2$ the exact solution

eu(h)	h = 1/24	h = 1/48	h = 1/96
L = 4	2.1e−04	2.6e−05	3.5e−06
L = 8	1.8e−04	2.3e−05	3.0e−06
L = 16	1.7e−04	2.2e−05	2.9e−06

ep(h)	h = 1/24	h = 1/48	h = 1/96
L = 4	6.7e−04	1.6e−04	4.0e−05
L = 8	6.8e−04	1.6e−04	4.2e−05
L = 16	6.8e−04	1.7e−04	4.3e−05

Table 2. Relative errors in velocity field and pressure for different values of h on $\Omega_L = [0,L] \times [0,1]$ and with the same configuration as in Table 1

$$u(x,y) = \begin{pmatrix} -\sin^3(\pi x)\sin^2(\pi y)\cos(\pi y) \\ -\sin^2(\pi x)\sin^3(\pi y)\cos(\pi x) \end{pmatrix}, \quad p(x,y) = (x-0.25)^2(y-0.25)^2$$

and partition Ω into 4 equal subdomains with a cross point at $(0.5,0.5)$. Table 3 shows the results and we see that the number of iterations is independent of the mesh size again (Fig. 1).

	Dual	Initial PCG	Final PCG		
h	# Iters	# Iters	# Iters	eu(h)	ep(h)
1/12	7	22	20	6.9e−4	4.2e−03
1/24	7	21	20	8.8e−5	1.0e−03
1/48	7	23	21	1.2e−5	2.5e−04
1/96	7	23	23	1.4e−6	8.3e−05

Table 3. Results obtained when subdividing the domain $\Omega = (0,1)^2$ into four subdomains with a cross point at $(0.5,0.5)$

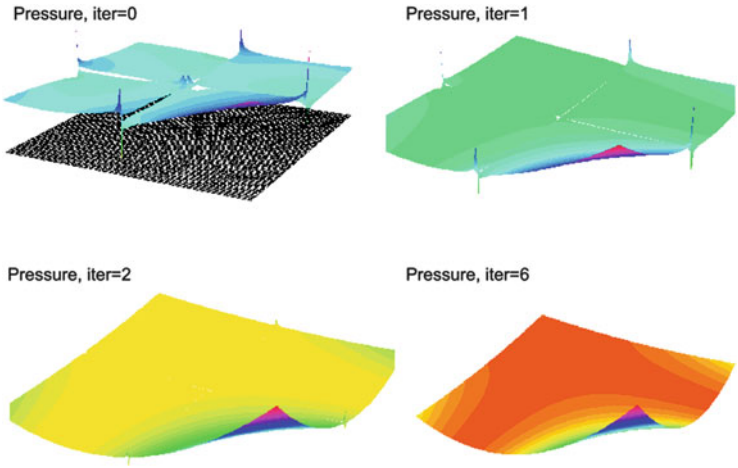


Fig. 1. Inital iteration with the underlying mesh and some contiguous iterations for the computed pressure

4 Conclusions

We presented a FETI-DP Mortar method applied to incompressible Stokes equations. Continuity at crosspoints is retained and the jumps across interfaces are included in the continuous formulation. The Lagrange multipliers are represented by their Riesz-canonical isometry, which improves their regularity from $H_{00}^{-1/2}(\Gamma)$ to $H_{00}^{1/2}(\Gamma)$, and the mortaring is performed using the $H_{00}^{1/2}(\Gamma)$ scalar product for each interface Γ. As a consequence, continuous bounds are replicated at the discrete level and no stabilization is required. In this setting we solve a dual problem by a CG that has a mesh independent condition number. The primal problems involved include the effect of the coupling between neighboring subdomains at interfaces and are solved by PCG. Still independent solves per subdomains are possible.

The advantage of the continuous framework introduced is the clear sight of the effect of condensing all information on subdomains and interfaces before the discrete work starts and the use of, to our belief, the most appropriated norms on subdomains and interfaces that make no necessary the use of mesh dependent norms for obtaining stability.

Acknowledgments Research partially funded by Spanish government MEC Research Project MTM2009-07719. The authors thanks Frédéric Nataf and Tomás Chacón Rebollo for many valuable comments on this work.

Bibliography

[1] Robert A. Adams. *Sobolev spaces*. Academic Press [A subsidiary of Harcourt Brace Jovanovich, Publishers], New York-London, 1975. Pure and Applied Mathematics, Vol. 65.

[2] C. Bernardi, T. Chacón Rebollo, and E. Chacón Vera. A FETI method with a mesh independent condition number for the iteration matrix. *Comput. Methods Appl. Mech. Engrg.*, 197(13-16):1410–1429, 2008. ISSN 0045-7825. doi: 10.1016/j.cma.2007.11.019. URL http://dx.doi.org/10.1016/j.cma.2007.11.019.

[3] Eliseo Chacón Vera. A continuous framework for FETI-DP with a mesh independent condition number for the dual problem. *Comput. Methods Appl. Mech. Engrg.*, 198(30-32):2470–2483, 2009. ISSN 0045-7825. doi: 10.1016/j.cma.2009.02.037. URL http://dx.doi.org/10.1016/j.cma.2009.02.037.

[4] Vivette Girault and Pierre-Arnaud Raviart. *Finite element methods for Navier-Stokes equations. Theory and algorithms.*, volume 5 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin, 1986. ISBN 3-540-15796-4.

[5] P. Grisvard. *Singularities in boundary value problems*, volume 22 of *Recherches en Mathématiques Appliquées [Research in Applied Mathematics]*. Masson, Paris, 1992. ISBN 2-225-82770-2.

One-Shot Domain Decomposition Methods for Shape Optimization Problems

Rongliang Chen[1] and Xiao-Chuan Cai[2]

[1] College of Mathematics and Econometrics, Hunan University, Changsha, Hunan 410082, China (rlchen@hnu.edu.cn)
[2] Department of Computer Science, University of Colorado at Boulder, Boulder, CO 80309, USA (cai@cs.colorado.edu)

1 Introduction

Shape optimization aims to optimize an objective function by changing the shape of the computational domain. In recent years, shape optimization has received considerable attentions. On the theoretical side there are several publications dealing with the existence of solution and the sensitivity analysis of the problem; see e.g., [6] and references therein. On the practical side, optimal shape design has played an important role in many industrial applications, for example, aerodynamic shape design [7], artery bypass design [1, 10], and so on. In this paper, we propose a general framework for the parallel solution of shape optimization problems, and study it in detail for the optimization of an artery bypass problem.

For PDE constrained optimization problems, there are two basic approaches: *nested analysis and design* and *simultaneous analysis and design* (one-shot methods). As computers become more powerful in processing speed and memory capacity, one-shot methods become more attractive due to their higher degree of parallelism, better scalability, and robustness in convergence. The main challenges in the one-shot approaches are that the nonlinear system is two to three times larger, and the corresponding indefinite Jacobian system is a lot more ill-conditioned and also much larger. So design a preconditioner that can substantially reduce the condition number of the large fully coupled system and, at the same time, provides the scalability for parallel computing becomes a very important stage in the one-shot methods. There are several recent publications on one-shot methods for PDE constrained optimization problems. In [5], a reduced Hessian sequential quadratic programming method was introduced for an aerodynamic design problem. In [4], a parallel *full space method* was introduced for the boundary control problem where a Newton-Krylov method is used together with Schur complement type preconditioners. In [9] and [8], an overlapping Schwarz based Lagrange-Newton-Krylov approach (LNKSz) was investigated for some boundary control problems. As far as we know no one has studied shape optimization problems using LNKSz, which has the potential to solve very large problems on machines with a large number of processors (*np*). The previ-

R. Bank et al. (eds.), *Domain Decomposition Methods in Science and Engineering XX*, Lecture Notes in Computational Science and Engineering 91,
DOI 10.1007/978-3-642-35275-1_63,

ous work on LNKSz doesn't consider the change of the computational domain which makes the study much more difficult and interesting.

2 Shape Optimization on a Moving Mesh

We consider a class of shape optimization problems governed by the stationary incompressible Navier-Stokes equations defined in a two dimensional domain Ω_α. Our goal is to computationally find the optimal shape for part of the boundary $\partial\Omega_\alpha$ such that a given objective function J_o is optimized. We represent the part of the boundary by a smooth function $\alpha(x)$ determined by a set of parameters $\mathbf{a} = (a_1, a_2, \ldots, a_p)$. By changing the shape defined by $\alpha(x)$, one can optimize certain properties of the flow. In this paper, we focus on the minimization of the energy dissipation in the whole flow field and use the integral of the squared energy deformation as the objective function [6]

$$\min_{\mathbf{u},\alpha} \; J_o(\mathbf{u},\alpha) = 2\mu \int_{\Omega_\alpha} \varepsilon(\mathbf{u}) : \varepsilon(\mathbf{u}) dxdy + \frac{\beta}{2} \int_I (\alpha'')^2 dx$$

subject to

$$\begin{cases} -\mu\Delta\mathbf{u} + \mathbf{u}\cdot\nabla\mathbf{u} + \nabla p = \mathbf{f} & \text{in} \quad \Omega_\alpha, \\ \nabla\cdot\mathbf{u} = 0 & \text{in} \quad \Omega_\alpha, \\ \mathbf{u} = \mathbf{g} & \text{on} \quad \Gamma_{inlet}, \\ \mathbf{u} = \mathbf{0} & \text{on} \quad \Gamma_{wall}, \\ \mu\dfrac{\partial\mathbf{u}}{\partial\mathbf{n}} - p\cdot\mathbf{n} = \mathbf{0} & \text{on} \quad \Gamma_{outlet}, \\ \alpha(a) = z_1, \quad \alpha(b) = z_2, \end{cases} \tag{1}$$

where $\mathbf{u} = (u, v)$ and p represent the velocity and pressure, $\mathbf{n}$ is the outward unit normal vector on $\partial\Omega_\alpha$ and μ is the kinematic viscosity. Γ_{inlet}, Γ_{outlet} and Γ_{wall} represent the inlet, outlet and wall boundaries, respectively; see Fig. 1. $\mathbf{f}$ is the given body force and $\mathbf{g}$ is the given velocity at the inlet Γ_{inlet}. $\varepsilon(\mathbf{u}) = \frac{1}{2}(\nabla\mathbf{u} + (\nabla\mathbf{u})^{\mathbf{T}})$ is the deformation tensor for the flow velocity $\mathbf{u}$ and β is a nonnegative constant. $I = [a, b]$ is an interval in which the shape function $\alpha(x)$ is defined. In the constraints, the first five equations are the Navier-Stokes equations and boundary conditions and the last two equations indicate that the optimized boundary should be connected to the rest of the boundary and z_1 and z_2 are two given constants. The last term in the objective function is a regularization term providing the regularity of $\partial\Omega_\alpha$.

The optimization problem (1) is discretized with a LBB-stable (*Ladyzhenskaya-Babuška-Brezzi*) $Q_2 - Q_1$ finite element method. Since the computational domain of the problem changes during the optimization process, the mesh needs to be modified following the computational domain. Generally speaking, there are two strategies to modify the mesh. One is mesh reconstruction which often guarantees a good new mesh but is computationally expensive. The other strategy is moving mesh which is cheaper but the deformed mesh may become ill-conditioned when the boundary variation is large. In our test case the boundary variations are not very large, so we

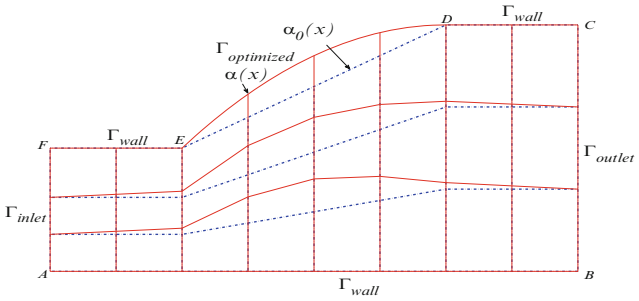


Fig. 1. The initial domain Ω_{α_0} (*dashed line*) and deformed domain Ω_α (*solid line*) over a simple mesh. The boundary $\Gamma_{optimized}$ (*ED*) denotes the part of the boundary whose shape is computed by the optimization process

use the latter strategy. The moving of the mesh is simply described by Laplace's equations.

$$\begin{cases} -\Delta\delta_{\mathbf{x}} = \mathbf{0} \text{ in } \Omega_{\alpha_0}, \\ \delta_{\mathbf{x}} = \mathbf{g}_\alpha \text{ on } \partial\Omega_{\alpha_0}, \end{cases} \tag{2}$$

where $\delta_{\mathbf{x}}$ is the mesh displacement and $\mathbf{g}_\alpha = (g_\alpha^x, g_\alpha^y)$ is the displacement on the boundary determined by $\alpha(x)$. Note that $\mathbf{g}_\alpha$ is obtained automatically during the iterative solution process. For example, in Fig. 1, $g_\alpha^x = 0$ and $g_\alpha^y = \alpha(x) - \alpha_0(x)$. The Eqs. (2) are discretized with a Q_2 finite element method. The discretized shape optimization problem is given as follows

$$\min_{\mathbf{u},\mathbf{a},\delta_{\mathbf{x}}} \quad J_o(\mathbf{u},\mathbf{a},\delta_{\mathbf{x}}) = \mu\mathbf{u}^{\mathbf{T}}\mathbf{J}\mathbf{u} + \frac{\beta}{2}\mathbf{J}_\alpha$$

subject to

$$\begin{cases} \mathbf{K}\mathbf{u} + \mathbf{B}(\mathbf{u})\mathbf{u} - \mathbf{Q}\mathbf{p} = \mathbf{F_f} + \mathbf{F_u}, \\ \mathbf{Q}^{\mathbf{T}}\mathbf{u} = \mathbf{0}, \\ \mathbf{D}\delta_{\mathbf{x}} = \mathbf{F_x}, \\ \mathbf{A_a} = \mathbf{F_a}. \end{cases} \tag{3}$$

Here $\mathbf{F_f}$ refers to the discretized body force, $\mathbf{F_u}$ and $\mathbf{F_x}$ refer to the Dirichlet boundary condition for $\mathbf{u}$ and $\delta_{\mathbf{x}}$, respectively, and $\mathbf{A_a}$ and $\mathbf{F_a}$ are the geometric constrains. Note that $\mathbf{K}$, $\mathbf{B}(\mathbf{u})$, $\mathbf{Q}$ and $\mathbf{J}$ depend on the grid displacement $\delta_{\mathbf{x}}$, while $\mathbf{D}$ is independent of $\delta_{\mathbf{x}}$. Here $\delta_{\mathbf{x}}$ is treated as an optimization variable and the moving mesh equations are viewed as constraints of the optimization problem which are solved simultaneously with the other equations.

3 One-Shot Lagrange-Newton-Krylov-Schwarz Methods

We use a Lagrange multiplier method to transform the optimization problem (3) to a nonlinear system $\mathbf{G}(\mathbf{X}) = \mathbf{0}$ which is solved by an inexact Newton method.

Given an initial guess $\mathbf{X^0}$, at each iteration, $k = 0, 1, \cdots$, we use a GMRES method to approximately solve the preconditioned system

$$\mathbf{H}^k(\mathbf{M}^k)^{-1}(\mathbf{M}^k\mathbf{d}^k) = -\mathbf{G}^k, \tag{4}$$

to find a search direction $\mathbf{d}^k$, where $\mathbf{H}^k = \nabla_X \mathbf{G}(\mathbf{X}^k)$ is the Jacobian matrix of the nonlinear function, $\mathbf{G}^k = \mathbf{G}(\mathbf{X}^k)$ and $(\mathbf{M}^k)^{-1}$ is an additive Schwarz preconditioner [11] defined as

$$(\mathbf{M}^k)^{-1} = \sum_{l=1}^{N_p} (R_l^\delta)^{\mathbf{T}} (\mathbf{H}_l^k)^{-1} R_l^\delta,$$

where $\mathbf{H}_l^k = R_l^\delta \ \mathbf{H}^k \ (R_l^\delta)^{\mathbf{T}}$, R_l^δ is a restriction operator from Ω_α to the overlapping subdomain, δ is the size of the overlap which is understood in terms of the number of elements; i.e., $\delta = 8$ means the overlapping size is 8 layers of elements, and N_p is the number of subdomains which is equal to np in this paper. After approximately solving (4), the new approximate solution is defined as $\mathbf{X}^{k+1} = \mathbf{X}^k + \tau^k \mathbf{d}^k$, and the step length τ^k is selected by a cubic line search.

4 Numerical Experiments

The algorithm introduced in the previous sections is applicable to general shape optimization problems governed by incompressible Navier-Stokes equations. Here we study an application of the algorithm for the incoming part of a simplified artery bypass problem[1] [2] as shown in Fig. 2. Our solver is implemented using PETSc [3]. All computations are performed on an IBM BlueGene/L supercomputer at the National Center for Atmospheric Research. Unstructured meshes, which are generated with CUBIT and partitioned with ParMETIS, are used in this paper.

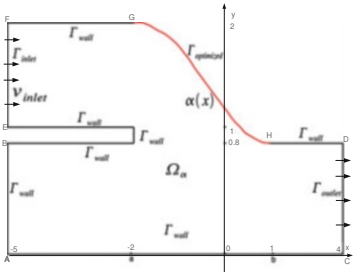


Fig. 2. The incoming part of a simplified bypass model; The *red* boundary $\Gamma_{optimized}$ denotes the part of the boundary whose shape is to be determined by the optimization process

[1] This is the incoming part of a bypass: www.reshealth.org/images/greystone/em\delimiter"026E30F_2405.gif

Without the blockage, the flow is supposed to go from AB to CD, but now we assume that AB is blocked and the flow has to go through EF. For simplicity, we let the thickness EF be fixed and the body force $\mathbf{f} = \mathbf{0}$ in the Navier-Stokes equations. The shape of the bypass is determined by the curve GH as in Fig. 2. The boundary conditions on the inlet Γ_{intlet} are chosen as a constant v_{in}, no-slip boundary conditions are used on the walls Γ_{wall}. On the outlet section Γ_{outlet}, the free-stress boundary conditions are imposed; see (1). We use a polynomial $\alpha(x) = \sum_{i=1}^{p} a_i x^i$ with $p = 7$ to represent the part of the boundary that needs to be optimized. Other shape functions can be used, but here we simply follow [1]. The goal is to compute the coefficients $\mathbf{a} = (a_1, \ldots, a_p)$, such that the energy loss is minimized.

In all experiments, we use a hand-coded Jacobian matrix. The Jacobian system in each Newton step is solved by a right-preconditioned restarted GMRES with an absolute tolerance of 10^{-10}, a relative tolerance of 10^{-3}, and a restart at 100. We stop the Newton iteration when the nonlinear residual is decreased by a factor of 10^{-6}.

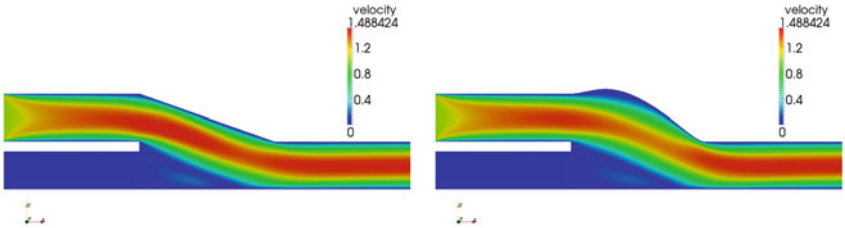


Fig. 3. Velocity distribution of the initial (*left*) and optimal shapes (*right*). The initial shape is given by a *straight line*. $\beta = 0.01$ and $Re = 100$

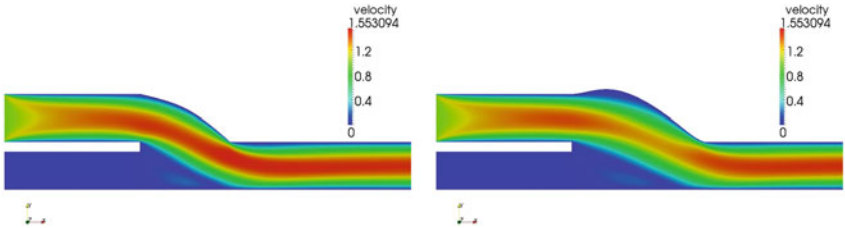


Fig. 4. Velocity distribution of the initial (*left*) and optimal shapes (*right*). The initial shape is given as $\alpha(x) = 0.4 + 0.45x^2 + 0.15x^3$. $\beta = 0.01$ and $Re = 100$

In the first test case, we set the Reynolds number $Re = \frac{Lv_{in}}{\mu}$ to 100, where $L = 1.0\,\text{cm}$ is the artery diameter, $v_{in} = 1.0\,\text{cm/s}$ is the inlet velocity and $\mu = 0.01\,\text{cm}^2/\text{s}$.

We solve the problem on a mesh with about 18,000 elements. $\beta = 0.01$ and the degrees of freedom (DOF) is 589,652. The initial shape is given by a straight line, and Fig. 3 shows the velocity distribution of the initial (left) and optimal shapes (right). The energy dissipation of the optimized shape is reduced by about 5.13 % compared to the initial shape. Figure 4 is the velocity distribution of another initial shape (left) which is given as $\alpha(x) = 0.4 + 0.45x^2 + 0.15x^3$ and the corresponding optimal shape (right). The reduction of the energy dissipation of this case is about 11.96 %. Figures 3 and 4 show that we can obtain nearly the same optimal shape from different initial shapes.

In the test case showed in Fig. 3, if we add a small inlet velocity at the boundary AB, which is equal to that the blood flow is not totally blocked, the computed optimal shape would be different from what is shown in Fig. 3. If we move the boundary AB towards CD (A from $(-5,0)$ to $(-3,0)$ and B from $(-5,0.8)$ to $(-3,0.8)$), the optimal shape is nearly the same as Fig. 3 since the flow in the "dead area" doesn't impact much of the optimal solution.

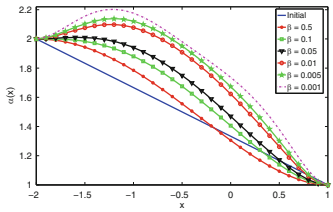


Fig. 5. The initial shape and optimal shapes with different values of parameter β. $DOF =$ 589,652 and $Re = 100$

The regularization parameter β in the objective function is very important for shape optimization problems. From Table 1 we see that reducing β can increase the reduction of the energy dissipation ("Init.", "Opt." and "Reduction" are the initial, optimized and reduction of the energy dissipation in the table), but the number of Newton (Newton) and the average number of GMRES iterations per Newton (GMRES) and the total compute time in seconds (Time) increase, which means that the nonlinear algebraic system is harder to solve when β is small. This is because the boundary of Ω_α is more flexible and may become irregular when β is too small. Figure 5 shows the initial shape and the optimized shapes obtained with different values of β. From this figure we see that β controls the boundary deformation.

To show the parallel scalability of the algorithm, two meshes with $DOF =$ 589,652 and $DOF = 928{,}572$ are considered. The strong scalability of our algorithm is good; see Fig. 6 and Table 2, which show that the speedup is almost linear when np is small. As expected in one-level Schwarz methods, the preconditioner becomes worse as the number of subdomains increases.

Table 3 shows some results for different Re. Judging from the increase of the number of linear and nonlinear iterations, it is clear that the problem becomes harder

Table 1. Effect of the parameter β. $DOF = 589,652$, $Re = 100$.

β	Newton	GMRES	Time	Energy Dissipation		
				Init.	Opt.	Reduction
0.05	4	386.00	477.89	1.17	1.12	4.27%
0.01	5	441.40	600.86	1.17	1.11	5.13%
0.005	5	439.00	599.77	1.17	1.10	5.98%
0.001	6	510.67	747.78	1.17	1.10	5.98%

Table 2. Parallel scalability for two different size grids. $\beta = 0.1$, $overlap = 6$ and $Re = 100$.

np	$DOF = 589,652$			$DOF = 928,572$		
	Newton	GMRES	Time	Newton	GMRES	Time
32	4	124.50	2959.73	—	——	——
64	4	179.25	980.48	4	146.50	2121.52
128	4	346.75	455.69	4	330.00	844.62
256	4	533.25	280.96	4	520.75	541.97
512	4	917.50	282.07	4	861.00	361.08

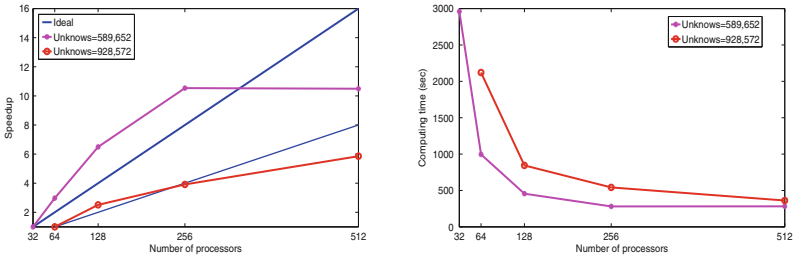


Fig. 6. The speedup and the total compute time for two different mesh sizes. $Re = 100$

as we increase the Re. On the other hand, we achieve higher percentage of reduction of energy dissipation in the harder to solve situations.

Table 3. The impact of Re. $\beta = 0.1$, $overlap = 8$, $DOF = 589,652$, $np = 128$.

Re	Newton	GMRES	Time	Energy Dissipation		
				Init.	Opt.	Reduction
100	4	346.75	456.83	1.17	1.13	3.42%
200	4	372.00	470.16	0.65	0.62	4.62%
300	6	671.00	871.19	12.56	11.80	6.05%
600	7	721.71	1035.84	7.43	6.97	6.19%

5 Conclusions and Future Work

We developed a parallel one-shot LNKSz for two-dimensional shape optimization problems governed by incompressible Navier-Stokes equations. We tested the algorithms for an artery bypass design problem with more than 900,000 DOF and up to 512 processors. The numerical results show that our method is quite robust with respect to the *Re* and the regularization parameter. The strong scalability is almost ideal when np is not too large. In the future, we plan to study some multilevel Schwarz methods which may improve the scalability when np is large.

Bibliography

[1] F. Abraham, M. Behr, and M. Heinkenschloss. Shape optimization in stationary blood flow: A numerical study of non-Newtonian effects. *Comput. Methods Biomech. Biomed. Engrg.*, 8:127–137, 2005.

[2] V. Agoshkov, A. Quarteroni, and G. Rozza. A mathematical approach in the design of arterial bypass using unsteady Stokes equations. *J. Sci. Comput.*, 28: 139–165, 2006.

[3] S. Balay, K. Buschelman, V. Eijkhout, W. D. Gropp, D. Kaushik, M. G. Knepley, L. C. McInnes, B. F. Smith, and H. Zhang. PETSc Users Manual. Technical report, Argonne National Laboratory, 2010.

[4] G. Biros and O. Ghattas. Parallel Lagrange-Newton-Krylov-Schur methods for PDE-constrained optimization. Part I: The Krylov-Schur solver. *SIAM J. Sci. Comput.*, 27:687–713, 2005.

[5] O. Ghattas and C. Orozco. A parallel reduced Hessian SQP method for shape optimization. In N. M. Alexandrov and M.Y. Hussaini, editors, *Multidisciplinary Design Optimization: State of the Art*, pages 133–152. SIAM, Philadelphia, 1997.

[6] M. D. Gunzburger. *Perspectives in Flow Control and Optimization: Advances in Design and Control*. SIAM, Philadelphia, 2003.

[7] B. Mohammadi and O. Pironneau. *Applied Shape Optimization for Fluids*. Oxford University Press, Oxford, 2001.

[8] E. Prudencio and X.-C. Cai. Parallel multilevel restricted Schwarz preconditioners with pollution removing for PDE-constrained optimization. *SIAM J. Sci. Comput.*, 29:964–985, 2007.

[9] E. Prudencio, R. Byrd, and X.-C. Cai. Parallel full space SQP Lagrange-Newton-Krylov-Schwarz algorithms for PDE-constrained optimization problems. *SIAM J. Sci. Comput.*, 27:1305–1328, 2006.

[10] A. Quarteroni and G. Rozza. Optimal control and shape optimization of aorto-coronaric bypass anastomoses. *Math. Models and Methods in Appl. Sci.*, 13: 1801–1823, 2003.

[11] A. Toselli and O. Widlund. *Domain Decomposition Methods: Algorithms and Theory*. Springer-Verlag, Berlin, 2005.

A Schur Complement Method for Compressible Navier-Stokes Equations

Thu-Huyen Dao[1,2], Michael Ndjinga[1], and Frédéric Magoulès[2]

[1] CEA–Saclay, DEN, DM2S, STMF, LMEC, F–91191 Gif–sur–Yvette, France
[2] Appl. Mat. and Syst. Lab., Ecole Centrale Paris, 92295 Châtenay–Malabry, France
thu-huyen.dao@cea.fr, michael.ndjinga@cea.fr,
frederic.magoules@hotmail.com

Summary. Domain decomposition methods were first developed for elliptic problems, taking advantage of the strong regularity of their solutions. In the last two decades, many investigations have been devoted to improve the performance of these methods for elliptic and parabolic problems. The situation is less clear for hyperbolic problems with possible singular solutions. In this paper, we will discuss a nonoverlapping domain decomposition method for nonlinear hyperbolic problems. We use the finite volume method and an implicit version of the Roe approximate Riemann solver, and propose a new interface variable inspired by Dolean and Lanteri [1]. The new variable makes the Schur complement approach simpler and allows the treatment of diffusion terms. Numerical results for the compressible Navier-Stokes equations in various 2D and 3D configurations such as the Sod shock tube problem or the lid driven cavity problem show that our method is robust and efficient. Comparisons of performances on parallel computers with up to 512 processors are also reported.

1 Introduction

When solving a nonlinear partial differential equation by an implicit scheme, one classically ends by solving a nonlinear algebraic system using a Newton method. At each step of this method we have to solve a linear system $\mathscr{A}(U^k)U^{k+1} = b(U^k)$. This task is computationally expensive in particular since the matrix $\mathscr{A}$ is usually non-symmetric and very ill-conditioned. It is therefore necessary to find an efficient preconditioner.

When the size of the system is large (as in the case of $3D$ computations), the parallel solution on multiple processors is essential to obtain reasonable computation times. Currently in the thermal hydraulic code, FLICA-OVAP (see [2]), the matrix $\mathscr{A}$ and the right hand side b are stored on multiple processors and the system is solved in parallel with a Krylov solver (classical incomplete factorization). Unfortunately, the parallel preconditioners of FLICA-OVAP only perform well on a few processors. In contrast, if we want to increase the number of processors these parallel preconditioners perform poorly. Tests were run on different test cases and led

R. Bank et al. (eds.), *Domain Decomposition Methods in Science and Engineering XX*, Lecture Notes in Computational Science and Engineering 91,
DOI 10.1007/978-3-642-35275-1_64,

us to conclude that it is often better not to use these parallel preconditioners, especially for $3D$ problems. This strategy does not make an optimal use of the available computational power. Hence we seek for more efficient methods to distribute the computations. We study and use a domain decomposition method as an alternative to the classical distribution.

The paper is organized as follows. In Sects. 2 and 3, we present the mathematical model and its numerical schemes. In Sect. 4, we first review the domain decomposition method proposed by Dolean and Lanteri [1] based on a Schwarz algorithm. We then introduce a new interface variable which makes the Schur complement approach simpler and allows for the treatment of diffusion terms. Section 5 presents a set of numerical experiments to validate our method, compares it with that of [1] concerning the robustness and efficiency and presents the scalability and the performance of different preconditioners.

2 Mathematical Model

The simplest model of FLICA-OVAP consists of the following three balance laws for the mass, the momentum and the energy:

$$\begin{cases} \frac{\partial \rho}{\partial t} & + \nabla \cdot \mathbf{q} & & = 0 \\ \frac{\partial \mathbf{q}}{\partial t} & + \nabla \cdot \left(\mathbf{q} \otimes \frac{\mathbf{q}}{\rho} + p \mathbb{I}_d \right) & - \nu \Delta (\frac{\mathbf{q}}{\rho}) & = 0 \\ \frac{\partial (\rho E)}{\partial t} & + \nabla \cdot \left[(\rho E + p) \frac{\mathbf{q}}{\rho} \right] & - \lambda \Delta T & = 0 \end{cases} \tag{1}$$

where ρ is the density, $\mathbf{v}$ the velocity, $\mathbf{q} = \rho \mathbf{v}$ the momentum, p the pressure, ρe the internal energy, $\rho E = \rho e + \frac{\|\mathbf{q}\|^2}{2\rho}$ the total energy, T the absolute temperature, ν the viscosity and λ the thermal conductivity. We close the system (1) by the ideal gas law $p = (\gamma - 1)\rho e$. For the sake of simplicity, we consider constant viscosity and conductivity, and neglect the contribution of viscous forces in the energy equation. By denoting $U = (\rho, \mathbf{q}, \rho E)^t$ the vector of conserved variables, the Navier–Stokes system (1) can be written as a nonlinear system of conservation laws:

$$\frac{\partial U}{\partial t} + \nabla \cdot (\mathscr{F}^{conv}(U)) + \nabla \cdot \left(\mathscr{F}^{diff}(U) \right) = 0, \tag{2}$$

where $\mathscr{F}^{conv}(U) = \begin{pmatrix} \mathbf{q} \\ \mathbf{q} \otimes \frac{\mathbf{q}}{\rho} + p \mathbb{I}_d \\ (\rho E + p) \frac{\mathbf{q}}{\rho} \end{pmatrix}$, $\mathscr{F}^{diff}(U) = \begin{pmatrix} 0 \\ -\nu \nabla (\frac{\mathbf{q}}{\rho}) \\ -\lambda \nabla T \end{pmatrix}$.

3 Numerical Method

The conservation form (2) allows for the definition of weak solutions, which can be discontinuous ones. Discontinuous solutions such as shock waves are of great

importance in transient calculations. In order to correctly capture shock waves, one needs a robust, low diffusive conservative scheme. The finite volume framework is the most appropriate setup to write discrete equations that express the conservation laws at each cell (see [3]).

We decompose the computational domain into N disjoint cells C_i with volume v_i. Two neighboring cells C_i and C_j have a common boundary ∂C_{ij} with area s_{ij}. We denote $N(i)$ the set of neighbors of a given cell C_i and $\mathbf{n}_{ij}$ the exterior unit normal vector of ∂C_{ij}. Integrating the system (2) over C_i and setting $U_i(t) = \frac{1}{v_i}\int_{C_i} U(x,t)dx$ and $U_i^n = U_i(n\Delta t)$, the discretized equations can be written:

$$\frac{U_i^{n+1}-U_i^n}{\Delta t} + \sum_{j\in N(i)} \frac{s_{ij}}{v_i}\left(\overrightarrow{\Phi}_{ij}^{conv} + \overrightarrow{\Phi}_{ij}^{diff}\right) = 0. \tag{3}$$

with: $\overrightarrow{\Phi}_{ij}^{conv} = \frac{1}{s_{ij}}\int_{\partial C_{ij}} \mathscr{F}^{conv}(U^{n+1})\cdot\mathbf{n}_{ij}ds$, $\overrightarrow{\Phi}_{ij}^{diff} = \frac{1}{s_{ij}}\int_{\partial C_{ij}} \mathscr{F}^{diff}(U^{n+1})\cdot\mathbf{n}_{ij}ds$.

To approximate the convection numerical flux $\overrightarrow{\Phi}_{ij}^{conv}$ we solve an approximate Riemann problem at the interface ∂C_{ij}. Using the Roe local linearisation of the fluxes [4], we obtain the following formula:

$$\overrightarrow{\Phi}_{ij}^{conv} = \frac{\mathscr{F}^{conv}(U_i^{n+1})+\mathscr{F}^{conv}(U_j^{n+1})}{2}\cdot\mathbf{n}_{ij} - \mathscr{D}(U_i^{n+1},U_j^{n+1})\frac{U_j^{n+1}-U_i^{n+1}}{2} \tag{4}$$

$$= \mathscr{F}^{conv}(U_i^{n+1})\mathbf{n}_{ij} + A^-(U_i^{n+1},U_j^{n+1})(U_j^{n+1}-U_i^{n+1}), \tag{5}$$

where $\mathscr{D}$ is an upwinding matrix, $A(U_i^{n+1},U_j^{n+1})$ the Roe matrix and $A^{\pm} = \frac{A\pm\mathscr{D}}{2}$. The choice $\mathscr{D}=0$ gives the centered scheme, whereas $\mathscr{D}=|A|$ gives the upwind scheme. For the Euler equations, we can build $A(U_i^{n+1},U_j^{n+1})$ explicitly using the Roe averaged state (see [3]).

The diffusion numerical flux $\overrightarrow{\Phi}_{ij}^{diff}$ is approximated on structured meshes using the formula:

$$\overrightarrow{\Phi}_{ij}^{diff} = D\left(\frac{U_i^{n+1}+U_j^{n+1}}{2}\right)(U_j^{n+1}-U_i^{n+1}) \tag{6}$$

with the matrix $D(U) = \begin{pmatrix} 0 & \mathbf{0} & 0 \\ \frac{\nu\mathbf{q}}{\rho^2} & \frac{-\nu}{\rho}\mathbb{I}_d & 0 \\ \frac{\lambda}{c_v}\left(\frac{c_vT}{\rho}-\frac{||\mathbf{q}||^2}{2\rho^3}\right) & \frac{\mathbf{q}^t\lambda}{\rho^2 c_v} & -\frac{\lambda}{c_v\rho} \end{pmatrix}$, where c_v is the heat capacity at constant volume.

3.1 Newton Scheme

Finally, since $\sum_{j\in N(i)}\mathscr{F}^{conv}(U_i^{n+1}).\mathbf{n}_{ij} = 0$, using (5) and (6) the Eq. (3) of the numerical scheme becomes:

$$\frac{U_i^{n+1}-U_i^n}{\Delta t} + \sum_{j\in N(i)}\frac{s_{ij}}{v_i}\{(A^- + D)(U_i^{n+1},U_j^{n+1})\}(U_j^{n+1}-U_i^{n+1}) = 0. \tag{7}$$

The system (7) is nonlinear, hence we use the following Newton iterative method to obtain the required solutions:

$$\frac{\delta U_i^{k+1}}{\Delta t} + \sum_{j\in N(i)} \frac{s_{ij}}{v_i}\left[(A^- + D)(U_i^k, U_j^k)\right]\left(\delta U_j^{k+1} - \delta U_i^{k+1}\right) = -\frac{U_i^k - U_i^n}{\Delta t} - \sum_{j\in N(i)} \frac{s_{ij}}{v_i}\left[(A^- + D)(U_i^k, U_j^k)\right](U_j^k - U_i^k), \tag{8}$$

where $\delta U_i^{k+1} = U_i^{k+1} - U_i^k$ is the variation of the k-th iterate that approximates the solution at time $n+1$.

4 Domain Decomposition Method

The principle of the domain decomposition method by Schur complement is to decompose the global problem into independent subproblems solved on each processor. More precisely, if we want to solve the problem:

$$\begin{cases} \frac{\partial U}{\partial t} + \nabla \cdot \mathscr{F}(U) = 0 \text{ in } \Omega \\ BU = g \qquad \text{on } \partial\Omega \end{cases} \tag{9}$$

on a partition of the original domain $\Omega = \cup_{I=1}^{K} \Omega_I$, defining U_I as the restriction of the solution U in the subdomain Ω_I, the algorithm of the domain decomposition method is then written as:

$$\begin{cases} \dfrac{\partial U_I}{\partial t} + \nabla \cdot \mathscr{F}(U_I) = 0 \text{ in } \Omega \\ BU_I = g \qquad \text{on } \partial\Omega \cap \partial\Omega_I \\ C_I U_I = C_I U_J \qquad \text{on } \partial\Omega_I \cap \partial\Omega_j \end{cases} \tag{10}$$

where C_I is an interface operator which we will clarify later.

4.1 Dolean and Lanteri Interface Variable

In the article [1], in order to make the subsystem (10) solution independent, Dolean et al introduced a redundant variable Φ_{ij}^{DL} at the domain interface between two cells i and j : $\Phi_{ij}^{DL} = A^+_{Roe,\mathbf{n}_{i,j}} U_i - A^-_{Roe,\mathbf{n}_{i,j}} U_j$ and then defined the orthogonal projectors $P^{\pm}$ on the eigenvectors subspaces such that

$$P^-(U_i,U_j)\delta\phi_{ij}^{Do} = A^-_{Roe,\mathbf{n}_{i,j}}\delta U_j^{k+1},\ P^+(U_i,U_j)\delta\phi_{ij}^{Do} = -A^+_{Roe,\mathbf{n}_{i,j}}\delta U_i^{k+1}$$

This strategy can only be applied to the Euler equations (Eq. (2) with no viscosity and heat conductivity terms) using the upwind scheme. In order to include diffusion terms in the model and to use various schemes, we introduce a new interface variable Φ_{ij} at the domain interface between two cells i and j:

$$\Phi_{ij} = U_j - U_i \tag{11}$$

4.2 A New Interface Variable

In the case where the cell i of the subdomain I is at the boundary and has to communicate with the neighboring subdomains, we can rewrite the system (8) as:

$$\begin{aligned}\frac{\delta U_i^{k+1}}{\Delta t} \quad &+ \quad \sum_{j\in I, j\in N(i)} \frac{s_{ij}}{v_i}\left[(A^- + D)(U_i^k, U_j^k)\right]\left(\delta U_j^{k+1} - \delta U_i^{k+1}\right) \\ &= \quad -\frac{U_i^k - U_i^n}{\Delta t} - \sum_{j\in N(i)} \frac{s_{ij}}{v_i}\left[(A^- + D)(U_i^k, U_j^k)\right](U_j^k - U_i^k) \\ &\quad - \sum_{j\notin I, j\in N(i)} \left[(A^- + D)(U_i^k, U_j^k)\right]\delta\phi_{ij}\end{aligned}$$

By defining $\mathscr{U}_I = (U_1, \ldots, U_m)^t$ the unknown vector of the subdomain I and

$$\delta\phi_{IJ} = (\delta\phi_{ij})_{i\in I, j\in J, j\in N(i)} \tag{12}$$

and by denoting $P = A^- + D$, we can write the linear system as:

$$\mathscr{A}(\mathscr{U}_I^k)\delta\mathscr{U}_I^{k+1} = b_I(\mathscr{U}^n, \mathscr{U}^k) - \sum_{J\in N(I)} P(\mathscr{U}_I^k, \mathscr{U}_J^k)\delta\phi_{IJ} \tag{13}$$

By taking into account Eqs. (11)–(13), we can build an extended system that distinguishes the internal unknowns from the interface ones:

$$\left(\begin{array}{cccc|c} \mathscr{A}_1 & 0 & \ldots & \ldots & P_1 \\ 0 & \mathscr{A}_2 & 0 & \ldots & P_2 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & \mathscr{A}_N & P_N \\ \hline M_1 & \ldots & \ldots & M_N & \mathbb{I} \end{array}\right)\left(\begin{array}{c} \delta\mathscr{U}_1 \\ \delta\mathscr{U}_2 \\ \ldots \\ \delta\mathscr{U}_N \\ \delta\Phi \end{array}\right) = \left(\begin{array}{c} b_1 \\ b_2 \\ \ldots \\ b_N \\ b_\phi \end{array}\right) \tag{14}$$

where $\mathscr{A}_I$ is the matrix that couples the unknowns associated with internal cells of Ω_I whereas M_I enables us to build $\delta\Phi$, the interface unknown on all coupling subdomain interfaces, from the δU_I. The internal unknowns can be eliminated in favor of the interface ones to yield the following interface system:

$$S\delta\phi = b_\phi \tag{15}$$

with

$$\begin{aligned}(S\delta\phi)_{IJ} &= \delta\phi_{IJ} + M_{IJ}\mathscr{A}_I^{-1}\sum_{K\in N(I)} P_{IK}\delta\phi_{IK} + M_{JI}\mathscr{A}_J^{-1}\sum_{K\in N(J)} P_{JK}\delta\phi_{JK} \\ (b_\phi)_{IJ} &= M_{IJ}{A_I}^{-1}b_I + M_{JI}{A_J}^{-1}b_J\end{aligned}$$

The Eq. (15) can be solved by, e.g., GMRES, BICGStab, or the Richardson methods.

5 Numerical Results

5.1 Validation

Figures 1 and 2 present the profile of the pressure after 10 time steps using the upwind scheme with CFL = 10 for the Euler equations. Our initial state is a pressurized ball at the center of a closed box and for $t > 0$ there are waves which propagate and reflect all over the box. The gas expands in the box and we can see the shock waves and the rarefaction waves. The solution is solved on a cartesian mesh of 200×200 cells.

Figures 3 and 4 show the streamlines of the steady state obtained using centered scheme to solve a lid driven cavity flow at Reynolds number 400 on a cartesian 50×50 mesh. The lid speed is 1 m/s, the maximum Mach number of the flow is 0.008. According to these results, we obtain the same solutions by using single or multiple domains.

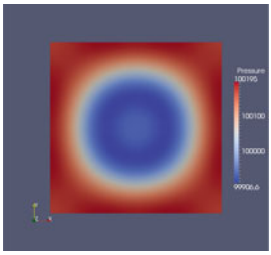


Fig. 1. Profile of the pressure at time step 10 on one processor

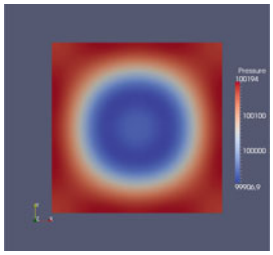


Fig. 2. Profile of the pressure at time step 10 on four processors

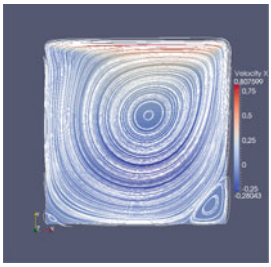


Fig. 3. Streamlines of V_x on one processor

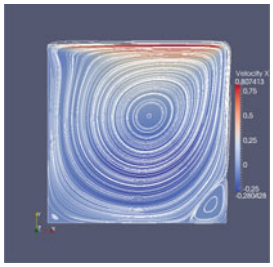


Fig. 4. Streamlines of V_x on four processors

5.2 Scalability

We now study the robustness and the scalability of our numerical method using the same test as presented in Sect. 5.1. In Figs. 5 and 6, we compare the parallel efficiency

of different preconditioners on 2*D* and 3*D* computations and with two and four processors. We see that without the preconditioner the solver is scalable. However, when

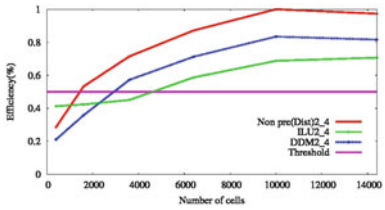


Fig. 5. Parallel efficiency for 2D Lid driven cavity

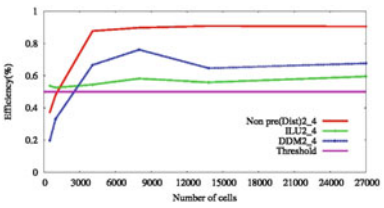


Fig. 6. Parallel efficiency for 3D Lid driven cavity

we use the Incomplete LU preconditioner, the scalability is not optimal especially for 3*D* problems. Our method proves better than ILU when we increase the number of cells in each subdomain. In Fig. 7, we compare the robustness of different methods

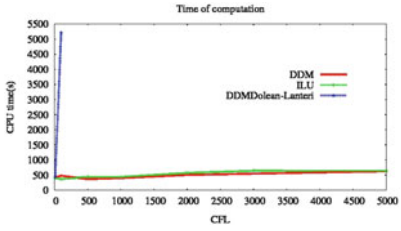


Fig. 7. Comparisons of parallelism in 3D Detonation, global mesh = $50 \times 50 \times 50$

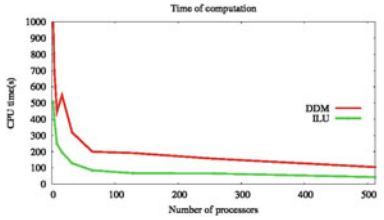


Fig. 8. Time of computation, 1 time step, global mesh = $96 \times 96 \times 96$

using the detonation problem. This problem is solved on a catersian $50 \times 50 \times 50$ cell mesh on two processors. The computation time of Dolean and Lanteri method increases rapidly because it needs many Newton iterations for convergence at each time step. In Fig. 8, we compare the scalability of the ILU preconditioner and of our method using the lid driven cavity problem solved on a global catersian $96 \times 96 \times 96$ cell mesh. The computation time of the domain decomposition method is higher than that of the ILU preconditioner due to the large number of Schur complement iterations.

6 Conclusion

We have presented a new interface variable which allows for the treatment of diffusion terms and the use of various numerical schemes. We also compared the efficiency and the scalability of our method with the classical distributed computations

and the method of Dolean and al. Our approach seems promising but we still need to find an efficient preconditioner for the Schur complement in order to reduce its computational time.

Bibliography

[1] V. Dolean and S. Lanteri. A domain decomposition approach to finite volume solution of the Euler equations on unstructured triangular meshes. *Int. J. Numer. Meth. Fluids*, 37(6), 2001.

[2] P. Fillion, A. Chanoine, S. Dellacherie, and A. Kumbaro. FLICA-OVAP: a new platform for core thermal-hydraulic studies. In *NURETH-13*, 2009.

[3] E. Godlewski and P.A. Raviart. *Numerical Approximation of Hyperbolic Systems of Conservation Laws*. Springer Verlag, 1996.

[4] P.L Roe. Approximate Riemann solvers, parameter vectors and difference schemes. *J. Comput. Phys.*, 43, 1981.

Numerical Study of the Almost Nested Case in a Multilevel Method Based on Non-nested Meshes

Thomas Dickopf and Rolf Krause

University of Lugano, Institute of Computational Science,
Via G. Buffi 13, 6904 Lugano, Switzerland
thomas.dickopf@usi.ch, rolf.krause@usi.ch

Summary. Partial differential equations in complex domains are very flexibly discretized by finite elements with unstructured meshes. For such problems, the challenging task to construct coarse level spaces for efficient multilevel preconditioners can in many cases be solved by a semi-geometric approach, which is based on a hierarchy of non-nested meshes. In this paper, we investigate the connection between the resulting semi-geometric multigrid methods and the truly geometric variant more closely. This is done by considering a sufficiently simple computational domain and treating the geometric multigrid method as a special case in a family of almost nested settings. We study perturbations of the meshes and analyze how efficiency and robustness depend on a truncation of the interlevel transfer. This gives a precise idea of which results can be achieved in the general unstructured case.

1 Introduction

This paper is about multilevel methods for an efficient solution of partial differential equations in complicated domains. Our particular purpose is to provide additional insight into the design of coarse spaces in case of unstructured finite element meshes. We study an approach of semi-geometric preconditioning based on non-nested mesh hierarchies motivated by Cai [2], Chan et al. [3, 4], Griebel and Schweitzer [6], Toselli and Widlund [8], and Xu [9]. This is a concept with rather weak requirements (yet still in a variational setting) compared with other geometry-based methods. The main contribution of the present paper is a numerical study of the almost nested case, which establishes a connection between the multilevel methods based on non-nested meshes and the standard variant. Combined with our investigations of mesh perturbations, this allows for the determination of a suitable truncation parameter for the interlevel transfer. As a result, the efficiency of the completely nested case is in large part retained.

R. Bank et al. (eds.), *Domain Decomposition Methods in Science and Engineering XX*, Lecture Notes in Computational Science and Engineering 91,
DOI 10.1007/978-3-642-35275-1_65,

2 Multilevel Preconditioners Based on Non-nested Meshes

This section aims at a semi-geometric preconditioning framework. We introduce a multiplicative multilevel preconditioner based on a hierarchy of non-nested meshes. This is done in a way which allows for a powerful convergence analysis as well as an efficient implementation.

Let $\Omega \subset \mathbb{R}^d$ be a Lipschitz domain of dimension $d \in \{2,3\}$. For a right hand side $\mathscr{F} \in H^{-1}(\Omega)$ and a positive function $\alpha \in L^\infty(\Omega)$ bounded away from zero, we consider the variational model problem

$$u \in H_0^1(\Omega): \quad a(u,v) := (\alpha \nabla u, \nabla v)_{L^2(\Omega)} = \mathscr{F}(v), \quad \forall\, v \in H_0^1(\Omega). \tag{1}$$

For a Galerkin discretization of problem (1), let $(\mathscr{T}_\ell)_{\ell\in\mathbb{N}}$ be a family of *non-nested* shape regular meshes of domains $(\Omega_\ell)_{\ell\in\mathbb{N}}$. We denote the set of nodes of $\mathscr{T}_\ell$ by $\mathscr{N}_\ell$ and abbreviate $n_\ell := |\mathscr{N}_\ell|$. At each level ℓ, we consider the space X_ℓ of Lagrange conforming finite elements of first order and denote its nodal basis as $\Lambda_\ell = (\lambda_p^\ell)_{p\in\mathscr{N}_\ell}$ with $\lambda_p^\ell(q) = \delta_{pq}$, $p,q \in \mathscr{N}_\ell$. For simplicity, we assume that $\Omega_L = \Omega$ and $X_L \subset H_0^1(\Omega)$ for a fixed finest level $L \geq 2$. In addition, let $\Omega_\ell \supset \Omega$ for all $\ell \in \{0,\dots,L-1\}$. The basic idea how the setting can be chosen is exemplarily illustrated in Fig. 1 (left) for an unstructured fine mesh with structured coarse meshes.

In the following, we consider an iterative method to efficiently solve the discrete problem, namely the ill-conditioned equation

$$\boldsymbol{A}_L \boldsymbol{u}_L = \boldsymbol{F}_L \quad \text{in } \mathbb{R}^{n_L}.$$

Here, $\boldsymbol{A}_L \in \mathbb{R}^{n_L\times n_L}$ is the stiffness matrix associated with X_L, i.e., $(\boldsymbol{A}_L)_{pq} := a(\lambda_p^L, \lambda_q^L)$ for $p,q \in \mathscr{N}_L$, and the right hand side $\boldsymbol{F}_L \in \mathbb{R}^{n_L}$ is given by $(\boldsymbol{F}_L)_p := \mathscr{F}(\lambda_p^L)$ for $p \in \mathscr{N}_L$.

For the construction of an appropriate coarse space hierarchy, let the spaces $(X_\ell)_{\ell=0,\dots,L}$ be connected by the prolongation operators $(\Pi_{\ell-1}^\ell)_{\ell=1,\dots,L}$, namely

$$\Pi_{\ell-1}^\ell : X_{\ell-1} \to X_\ell, \quad \forall\, \ell \in \{1,\dots,L\}.$$

The choice of a concrete transfer concept generating a set of suitable linear operators $(\Pi_{\ell-1}^\ell)_{\ell=1,\dots,L}$ in practice is discussed in full detail in [5]. An example is nodal interpolation. Now, let $V_L := X_L$; we emphasize that the fine space will not be touched in the present framework. We construct a nested sequence of spaces $(V_\ell)_{\ell=0,\dots,L}$ via

$$V_\ell := \Pi_{L-1}^L \cdots \Pi_\ell^{\ell+1} X_\ell, \quad \forall\, \ell \in \{0,\dots,L-1\}.$$

The images of the compositions of the given operators determine the coarse spaces.

With the nodal bases $(\Lambda_\ell)_{\ell=0,\dots,L}$, matrix representations $\boldsymbol{\Pi}_{\ell-1}^\ell \in \mathbb{R}^{n_\ell\times n_{\ell-1}}$ of $\Pi_{\ell-1}^\ell$ can be computed for $\ell \in \{1,\dots,L\}$ via $\boldsymbol{\Pi}_{\ell-1}^\ell \boldsymbol{v} := \boldsymbol{\Phi}_\ell^{-1}(\Pi_{\ell-1}^\ell \boldsymbol{\Phi}_{\ell-1}(\boldsymbol{v}))$ for all $\boldsymbol{v} \in \mathbb{R}^{n_{\ell-1}}$ with the coordinate isomorphisms $\boldsymbol{\Phi}_\ell : \mathbb{R}^{n_\ell} \to X_\ell$. Assume that these matrices have full rank. Then, bases of $(V_\ell)_{\ell=0,\dots,L-1}$ can recursively be defined by

$$\widetilde{\lambda}_q^\ell := \sum_{p\in\mathcal{N}_{\ell+1}} (\boldsymbol{\Pi}_\ell^{\ell+1})_{pq}\widetilde{\lambda}_p^{\ell+1}, \quad \forall\, q\in\mathcal{N}_\ell,$$

starting with $\widetilde{\lambda}_q^L := \lambda_q^L$ for $q\in\mathcal{N}_L$. The new coordinate isomorphisms with respect to the bases $\widetilde{\Lambda}_\ell := (\widetilde{\lambda}_p^\ell)_{p\in\mathcal{N}_\ell}$, $\ell\in\{0,\dots,L\}$, will be denoted by $\widetilde{\boldsymbol{\Phi}}_\ell : \mathbb{R}^{n_\ell}\to V_\ell$. Moreover, $\boldsymbol{M}_\ell\in\mathbb{R}^{n_\ell\times n_\ell}$ is the mass matrix with respect to $\widetilde{\Lambda}_\ell$, i.e., $(\boldsymbol{M}_\ell)_{pq} := (\widetilde{\lambda}_p^\ell,\widetilde{\lambda}_q^\ell)_{L^2(\Omega)}$ for $p,q\in\mathcal{N}_\ell$, $\ell\in\{0,\dots,L\}$.

Note that the mapping $\Pi_{\ell-1}^\ell$ between the given spaces $X_{\ell-1}$ and X_ℓ usually does not act on $V_{\ell-1}$ directly. Still, the matrix $\boldsymbol{\Pi}_{\ell-1}^\ell$ determines a linear transfer operator $\widetilde{\Pi}_{\ell-1}^\ell : V_{\ell-1}\to V_\ell$ by

$$v\mapsto \widetilde{\Pi}_{\ell-1}^\ell v := \widetilde{\boldsymbol{\Phi}}_\ell(\boldsymbol{\Pi}_{\ell-1}^\ell\widetilde{\boldsymbol{\Phi}}_{\ell-1}^{-1}(v)), \quad \forall\, v\in V_{\ell-1}, \quad \forall\,\ell\in\{1,\dots,L\}.$$

One can easily see that $\widetilde{\Pi}_{\ell-1}^\ell$ is the natural embedding because it interpolates the respective basis exactly. Thus, we can regard the matrix $\boldsymbol{\Pi}_{\ell-1}^\ell$ as an algebraic representation of the natural embedding of $V_{\ell-1}$ into V_ℓ. Consequently, the L^2-projection from V_ℓ to $V_{\ell-1}$ is represented by the matrix $\boldsymbol{M}_{\ell-1}^{-1}(\boldsymbol{\Pi}_{\ell-1}^\ell)^T\boldsymbol{M}_\ell\in\mathbb{R}^{n_{\ell-1}\times n_\ell}$. This holds true for any imaginable set of operators between the original non-nested spaces $(X_\ell)_{\ell=0,\dots,L}$; no special structure is required.

With this information we can summarize our efforts as follows. From the completely unrelated finite element spaces $(X_\ell)_{\ell=0,\dots,L}$ we have constructed a sequence of nested spaces $(V_\ell)_{\ell=0,\dots,L}$ such that the given prolongation operators $(\Pi_{\ell-1}^\ell)_{\ell=1,\dots,L}$ induce the natural embeddings $(V_{\ell-1}\hookrightarrow V_\ell)_{\ell=1,\dots,L}$ by their matrix representations $(\boldsymbol{\Pi}_{\ell-1}^\ell)_{\ell=1,\dots,L}$ with respect to the original bases $(\Lambda_\ell)_{\ell=0,\dots,L}$. In particular, the coarse level matrices for the nested spaces with the respective bases $\widetilde{\Lambda}_\ell$, as customary in a variational approach, can be written as

$$\boldsymbol{A}_{\ell-1} = (\boldsymbol{\Pi}_{\ell-1}^\ell)^T\boldsymbol{A}_\ell\,\boldsymbol{\Pi}_{\ell-1}^\ell\in\mathbb{R}^{n_{\ell-1}\times n_{\ell-1}}, \quad \forall\,\ell\in\{1,\dots,L\}. \tag{2}$$

If $\boldsymbol{A}_L$ is symmetric positive definite and if $\boldsymbol{\Pi}_{\ell-1}^\ell$ has full rank for all $\ell\in\{1,\dots,L\}$, the respective coarse level matrices $(\boldsymbol{A}_\ell)_{\ell=0,\dots,L-1}$ are symmetric positive definite, too. Note that the bandwidth of the coarse matrices depends on the transfer concept employed to obtain the prolongation operators.

The multiplicative Schwarz method studied in this paper is the symmetric multigrid $\mathcal{V}$-cycle in the novel space hierarchy $(V_\ell)_{\ell=0,\dots,L}$, which combines (Gauß–Seidel) smoothing and coarse level correction in the standard way. Naturally, only multiplications with the matrices $(\boldsymbol{\Pi}_{\ell-1}^\ell)_{\ell=1,\dots,L}$ and their transposes appear in the interlevel transfer of the algorithm; no mass matrices need to be inverted. Given the meshes $(\mathcal{T}_\ell)_{\ell=0,\dots,L}$ and a suitable transfer concept, we can compute all auxiliary matrices in a setup phase.

For a complete convergence analysis of this class of algorithms, which puts the semi-geometric approach into the well-known context of [1], we refer to [5]. There, we carefully distinguish between the generally different domains $(\Omega_\ell)_{\ell=0,\dots,L}$ and elaborate requirements for the meshes and the interlevel transfer to obtain a quasi-optimal result.

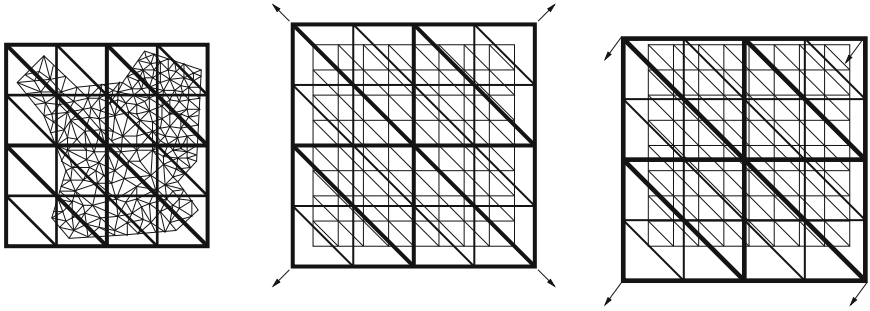


Fig. 1. Simplified sketch in $d = 2$. Basic idea of the coarse space construction based on non-nested meshes (*left*). Concerning the experiments: scaling (*center*) and translation (*right*) of the coarse meshes keeping the respective fine mesh fixed. We emphasize that all computations are in $d = 3$

The geometric nature of the construction usually requires some modifications of the meshes and operators, e.g., to ensure full rank. Moreover, a prevalent technique to keep the operator complexity $\mathscr{C}_{\mathrm{op}} := \sum_{\ell=0}^{L} n_\ell^A / n_L^A$ small, where n_ℓ^A is the number of non-zero entries of $\boldsymbol{A}_\ell$, is truncation of the prolongation operators by deleting the entries of $(\boldsymbol{\Pi}_{\ell-1}^{\ell})_{\ell=1,\dots,L}$ which are less than a truncation parameter $\varepsilon_{\mathrm{tr}} > 0$ times the maximal entry in the respective row. Afterwards, the modified rows are rescaled such that the row totals remain unchanged; see [7]. All this is done in the setup before the computation of the respective Galerkin products (2). In this paper, we choose $\Pi_{\ell-1}^{\ell}$ as standard nodal interpolation in X_ℓ for $\ell \in \{1,\dots,L\}$, namely $\Pi_{\ell-1}^{\ell} v := \sum_{p \in \mathscr{N}_\ell} v(p) \lambda_p^\ell$ for all $v \in X_{\ell-1}$, and refer to [5] for a detailed discussion.

3 Numerical Studies

3.1 The Almost Nested Limiting Case

We consider a hierarchy of four nested meshes $(\mathscr{T}_\ell)_{\ell=0,\dots,3}$ of the unit cube in $\mathbb{R}^3$ where the coarsest mesh consists of 768 elements with 189 nodes. Throughout the study, we keep the finest mesh $\mathscr{T}_L = \mathscr{T}_3$ with 393,216 elements and 68,705 nodes fixed. In contrast, the coarse domains $(\Omega_\ell)_{\ell<3}$ and the corresponding coarse meshes $(\mathscr{T}_\ell)_{\ell<3}$ are scaled around the center with a different factor between 0.95 and 1.05 for each set of tests; see Fig. 1 (center).

In the semi-geometric framework, it is absolutely necessary to perform a truncation procedure to retain the optimality of the algorithms. Otherwise, one can in general not prevent the appearance of very small and thus irrelevant entries in the prolongation matrices. We study the complexity of the constructed space hierarchy and the convergence of the semi-geometric multigrid method (stand-alone or in a preconditioned conjugate gradient method) for a variety of values for the parameter $\varepsilon_{\mathrm{tr}}$ in $[0.01, 0.49]$. Note that, for linear finite elements associated with simplicial meshes, it does generally not make sense to choose $\varepsilon_{\mathrm{tr}}$ greater than or equal to 0.5.

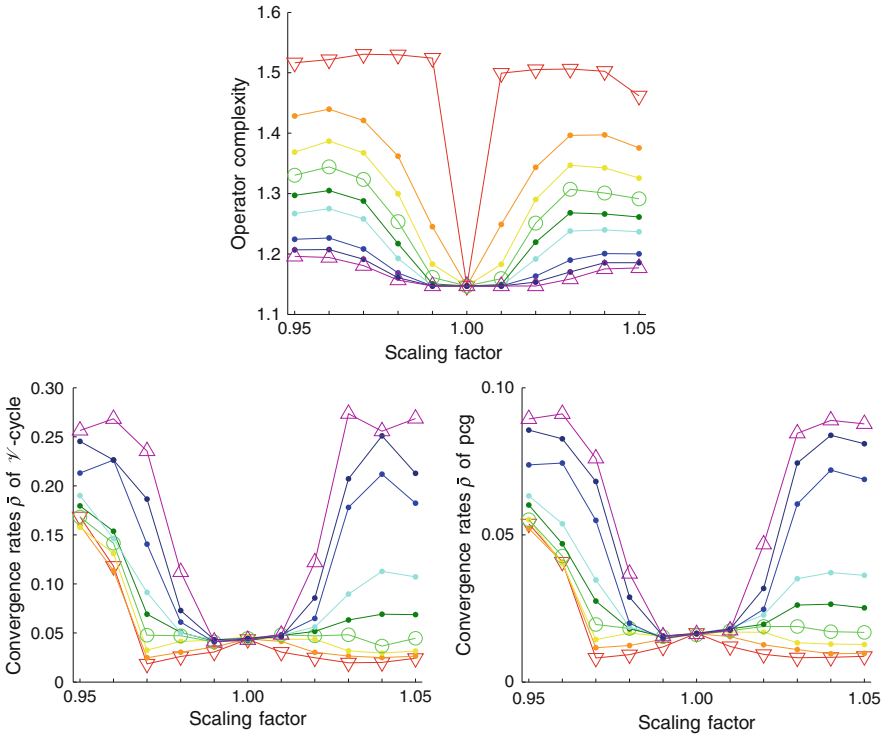


Fig. 2. The complexity measure $\mathscr{C}_{\text{op}}$ (*top*) and the convergence rates $\bar{\rho}_{\mathscr{V}(2,2)}$ (*left*) and $\bar{\rho}^{\text{pcg}}_{\mathscr{V}(2,2)}$ (*right*) of a semi-geometric multigrid method, plotted versus the scale of the coarse meshes. Each line represents a different parameter $\varepsilon_{\text{tr}} \in [0.01, 0.49]$. The *marked lines* correspond to the values 0.01 (▽), 0.20 (○) and 0.49 (△), respectively

This is because such a choice would result in deleting entries even in case of perfectly nested meshes, leaving nodes without direct coupling to the next coarser level.

The results of the experiments with scaled $(\Omega_\ell)_{\ell<3}$ are illustrated in Fig. 2. Each single line represents either the complexity $\mathscr{C}_{\text{op}}$ or one of the asymptotic convergence rates $\bar{\rho}_{\mathscr{V}(2,2)}$ and $\bar{\rho}^{\text{pcg}}_{\mathscr{V}(2,2)}$ for a fixed parameter ε_{tr} plotted versus the scale of the coarse meshes. The lines corresponding to the extreme ε_{tr}-values 0.01 and 0.49 are marked by downward and upward triangles, respectively; an intermediate value of 0.20 is marked by circles. Table 1 contains the numbers for these three values. We stop with the scales 0.95 and 1.05, respectively. For smaller factors, the convergence rates further increase quite fast as less and less of the computational domain $\Omega = \Omega_L$ is covered by the coarse meshes; the complexity measures do not change much in this case. For larger factors, the convergence rates slowly increase whereas the complexity measures decrease. This is due to the fact that more and more elements of the coarse meshes lie completely outside the computational domain.

scale	$\mathscr{C}_{\mathrm{op}}$	$\bar{\rho}_{\mathscr{V}(2,2)}$	$\bar{\rho}^{\mathrm{pcg}}_{\mathscr{V}(2,2)}$	$\mathscr{C}_{\mathrm{op}}$	$\bar{\rho}_{\mathscr{V}(2,2)}$	$\bar{\rho}^{\mathrm{pcg}}_{\mathscr{V}(2,2)}$	$\mathscr{C}_{\mathrm{op}}$	$\bar{\rho}_{\mathscr{V}(2,2)}$	$\bar{\rho}^{\mathrm{pcg}}_{\mathscr{V}(2,2)}$
0.95	1.52	0.169	0.054	1.33	0.168	0.055	1.20	0.256	0.089
0.96	1.52	0.118	0.041	1.34	0.142	0.043	1.19	0.268	0.091
0.97	1.53	0.018	0.008	1.32	0.048	0.020	1.18	0.235	0.076
0.98	1.53	0.026	0.009	1.25	0.047	0.018	1.16	0.112	0.037
0.99	1.52	0.031	0.012	1.16	0.041	0.015	1.15	0.041	0.016
1.00	1.15	0.044	0.016	1.15	0.044	0.016	1.15	0.044	0.016
1.01	1.50	0.031	0.012	1.16	0.048	0.017	1.15	0.048	0.018
1.02	1.51	0.025	0.009	1.25	0.047	0.019	1.15	0.122	0.047
1.03	1.51	0.020	0.008	1.31	0.048	0.019	1.16	0.273	0.085
1.04	1.50	0.020	0.008	1.30	0.037	0.017	1.18	0.256	0.089
1.05	1.46	0.024	0.009	1.29	0.045	0.017	1.18	0.269	0.088
	$\varepsilon_{\mathrm{tr}} = 0.01$			$\varepsilon_{\mathrm{tr}} = 0.20$			$\varepsilon_{\mathrm{tr}} = 0.49$		

Table 1. Studying the convergence behavior for a family of almost nested meshes associated with the unit cube. The middle row (scale 1.00) corresponds to the completely nested case in which the approach coincides with the standard geometric multigrid method

3.2 Robustness of the Coarse Level Hierarchy

The second experiment is to further investigate the influence of perturbations of the meshes on the coarse level hierarchy and the multigrid performance. Here, we consider different translations of the coarse meshes associated with the cube of scale 1.05 in direction of the unit vector $(\frac{2}{3}, \frac{2}{3}, \frac{1}{3})^T \in \mathbb{R}^3$ by sizes up to 0.12. In this case, the computational domain $\Omega = \Omega_L$ is covered by the domains $(\Omega_\ell)_{\ell<L}$ for almost the entire range of translations; see Fig. 1 (right). Basic robustness of the semi-geometric construction is demonstrated by the results in Fig. 3 where the parameter $\varepsilon_{\mathrm{tr}}$ again varies in the interval $[0.01, 0.49]$.

4 Discussion of the Results

As expected and observed in the vast majority of experiments, the convergence rates principally increase with increasing truncation parameter, which indicates that the constructed coarse spaces have adequate approximation power. Note that the deterioration of the convergence behavior is usually rather slow, though. It is evident that the semi-geometric methods, which leave the coarse meshes flexible, coincide with the standard geometric variants in the special case of nested meshes. In addition, an important observation from Sect. 3.1 is that both the complexities $\mathscr{C}_{\mathrm{op}}$ and the convergence rates of the geometric multigrid methods are retained in case the meshes are almost nested if a suitable parameter $\varepsilon_{\mathrm{tr}}$ is applied; see the discussion below. This also indicates that our construction is robust in the sense that the coarse level hierarchy (and with it the multigrid convergence) only varies slightly if the coarse meshes themselves change slightly. Perturbations of the meshes are irrelevant for the

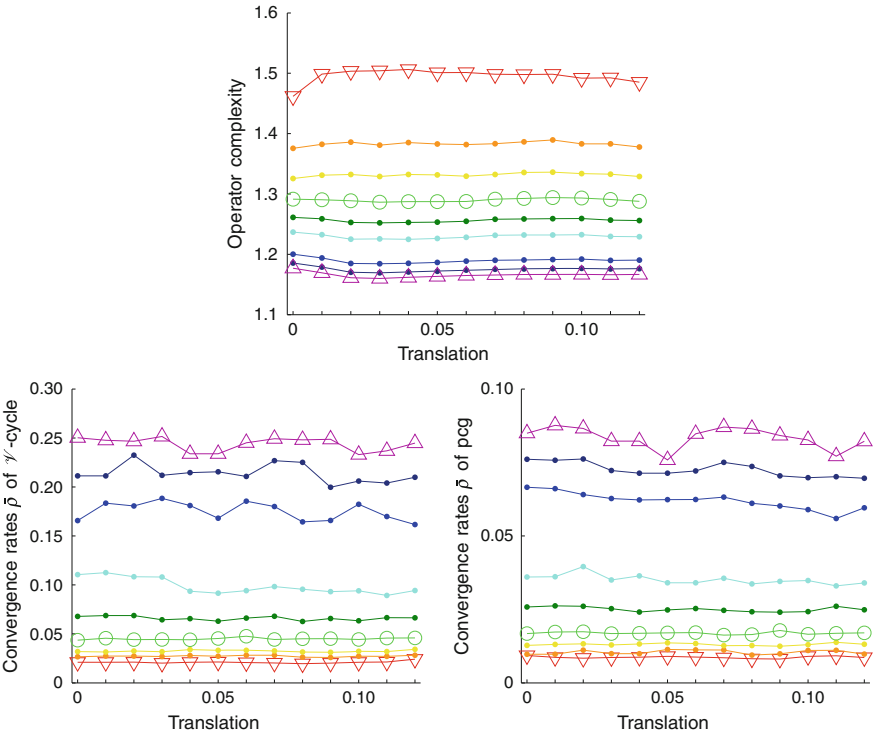


Fig. 3. The numbers $\mathscr{C}_{\text{op}}$ (*top*), $\bar{\rho}_{\mathscr{V}(2,2)}$ (*left*), and $\bar{\rho}^{\text{pcg}}_{\mathscr{V}(2,2)}$ (*right*). Each line represents a different parameter $\varepsilon_{\text{tr}} \in [0.01, 0.49]$ plotted versus the size of the coarse mesh translation

efficiency of the methods. This can also be seen clearly in the experiments described in Sect. 3.2.

As a general rule, we observe the following effects in Sect. 3.1. The larger the parameter ε_{tr} the less sensitive is the complexity $\mathscr{C}_{\text{op}}$ to changes of the coarse meshes. The smaller ε_{tr} the less sensitive are the convergence rates to changes of the coarse meshes. In our examples, the convergence actually improves in case of small perturbations for sufficiently small ε_{tr}. This is of course accompanied by a rapid increase of $\mathscr{C}_{\text{op}}$. The choice $\varepsilon_{\text{tr}} = 0.20$ (which is, interestingly enough, a standard value in many algebraic multigrid algorithms) is a reasonable attempt to achieve the two competing goals. It manages to keep the convergence rates almost constant for a rather broad range of different problem sizes while leading to an only moderate increase of $\mathscr{C}_{\text{op}}$.

Finally, let us compare to the general semi-geometric case. For an unstructured mesh with similar size (64,833 nodes) approximating a ball, the measured rates, $\bar{\rho}_{\mathscr{V}(2,2)} = 0.060$ and $\bar{\rho}^{\text{pcg}}_{\mathscr{V}(2,2)} = 0.024$, are not much worse than the ones produced by the geometric method on the cube with completely nested meshes, $\bar{\rho}_{\mathscr{V}(2,2)} = 0.044$ and $\bar{\rho}^{\text{pcg}}_{\mathscr{V}(2,2)} = 0.016$. However, for unstructured meshes without natural coarse level hierarchy, it seems impossible to achieve this fast convergence with an operator complexity as small as 1.15 which is easily obtained in the structured case.

For comparison, we have $\mathscr{C}_{\mathrm{op}} = 1.38$ for the ball. A whole series of experiments studying the asymptotics of the semi-geometric preconditioners can be found in [5].

5 Conclusion

In this paper, we reported on numerical studies of a class of preconditioners based on non-nested meshes. Considering the almost nested case, we determined a truncation parameter $\varepsilon_{\mathrm{tr}} = 0.20$ of the interlevel transfer to be reasonable in order to ensure that the efficiency of the completely nested case is in large part retained. Moreover, perturbations of the meshes turned out to be irrelevant for the efficiency of the methods.

Our results also show that, in the variational coarse space construction, it is appropriate to choose auxiliary meshes mimicking geometric coarsening, which leads to particularly small hierarchical overhead (less than 40 %). This is in contrast to the non-variational variant of the auxiliary space method [9] where both analysis and experiments indicate that the sizes of the original space and of the auxiliary space need to be comparable in a quite restrictive sense such that $\mathscr{C}_{\mathrm{op}}$ is usually clearly larger than two.

Acknowledgments This work was supported by the Bonn International Graduate School in Mathematics and by the Iniziativa Ticino in Rete.

Bibliography

[1] J.H. Bramble, J.E. Pasciak, J. Wang, and J. Xu. Convergence estimates for multigrid algorithms without regularity assumptions. *Math. Comput.*, 57(195):23–45, 1991.

[2] X. Cai. The use of pointwise interpolation in domain decomposition methods with non-nested meshes. *SIAM J. Sci. Comput.*, 16(1):250–256, 1995.

[3] T. Chan, B. Smith, and J. Zou. Overlapping Schwarz methods on unstructured meshes using non-matching coarse grids. *Numer. Math.*, 73(2):149–167, 1996.

[4] T. Chan, J. Xu, and L. Zikatanov. An agglomeration multigrid method for unstructured grids. In J. Mandel et al., editor, *Domain Decomposition Methods 10*, volume 218 of *Contemp. Math.*, pages 67–81. AMS: Providence, RI, 1998.

[5] T. Dickopf. *Multilevel Methods Based on Non-Nested Meshes*. PhD thesis, University of Bonn, 2010. http://hss.ulb.uni-bonn.de/2010/2365.

[6] M. Griebel and M.A. Schweitzer. A particle-partition of unity method. Part III: A multilevel solver. *SIAM J. Sci. Comput.*, 24(2):377–409, 2002.

[7] K. Stüben. An introduction to algebraic multigrid. In U. Trottenberg et al., editor, *Multigrid*, pages 413–532. Academic Press, London, 2001.

[8] A. Toselli and O. Widlund. *Domain Decomposition Methods – Algorithms and Theory*, volume 34 of *Springer Ser. Comput. Math.* Springer, 2005.

[9] J. Xu. The auxiliary space method and optimal multigrid preconditioning techniques for unstructured grids. *Computing*, 56(3):215–235, 1996.

BDDC for Higher-Order Discontinuous Galerkin Discretizations

Laslo Diosady[1] and David Darmofal[2]

[1] Massachusetts Institute of Technology diosady@alum.mit.edu
[2] Massachusetts Institute of Technology darmofal@mit.edu

Summary. The BDDC algorithm is extended to a large class of discontinuous Galerkin (DG) discretizations of second order elliptic problems in two spatial dimensions. An estimate of $C(1+\log(p^2H/h))^2$ is obtained for the condition number of the preconditioned system where C is a constant independent of p, h or H. Numerical simulations are presented which confirm the theoretical results

1 Introduction

A Balancing Domain Decomposition by Constraints (BDDC) method is presented for the solution of a discontinuous Galerkin (DG) discretization of a second-order elliptic problem in two dimensions. BDDC was originally introduced in [8] for the solution of continuous finite element discretizations. Mandel and Dohrmann [13] later proved a condition number bound of $\kappa \leq C(1+\log(H/h))^2$ for preconditioned system of a continuous finite element discretization of second order elliptic problems. Pavarino [15] and Klawonn et al. [11] extended the BDDC algorithm to higher-order finite element methods and proved a condition number bound of $\kappa \leq C(1+\log(p^2H/h))^2$. Further analysis of BDDC methods and their connection to FETI methods has been presented in [12, 14].

While domain decomposition methods have been widely studied for continuous finite element discretizations, relatively little work has been performed for discontinuous Galerkin discretizations. Previous work on domain decomposition methods for DG discretizations include [1, 10] and [9]. This work presents a BDDC method applied to a large class of DG methods considered in the unified analysis of [2]. A key component for the development and analysis of the BDDC algorithm involves presenting the DG discretization as the sum of element-wise "local" bilinear forms. The element-wise perspective leads naturally to the appropriate choice for the subdomain-wise local bilinear forms. Additionally, this perspective enables a connection to be drawn between the DG discretization and a related continuous finite element discretization. As a result of this connection, the condition number bound

R. Bank et al. (eds.), *Domain Decomposition Methods in Science and Engineering XX*, Lecture Notes in Computational Science and Engineering 91,
DOI 10.1007/978-3-642-35275-1_66,

for the BDDC preconditioned system for a large class of conservative and consistent DG methods is identical to that for continuous finite element methods.

2 DG Discretization

Consider the second order elliptic equation in a domain $\Omega \subset \mathscr{R}^2$:

$$-\nabla \cdot (\rho \nabla u) = f \qquad \text{in } \Omega, \qquad\qquad u = 0 \qquad \text{on } \partial\Omega \tag{1}$$

with positive $\rho > 0 \in L^\infty(\Omega)$, $f \in L^2(\Omega)$. Let the triangulation $\mathscr{T}$ be a partition of Ω into triangles or quadrilaterals. In order to simplify the presentation we assume that ρ takes on a constant value, ρ_κ on each element κ. Define $\mathscr{E}$ to be the union of edges of elements κ. Additionally, define $\mathscr{E}^i \subset \mathscr{E}$ and $\mathscr{E}^\partial \subset \mathscr{E}$ to be the set of interior, respectively boundary edges. Note that any edge $e \in \mathscr{E}^i$ is shared by two adjacent elements κ^+ and κ^- with corresponding outward pointing normal vectors $\boldsymbol{n}^+$ and $\boldsymbol{n}^-$. Let $\mathscr{P}^p(\kappa)$ denote the space of polynomials of order at most p on κ and define the following finite element space $W_h^p := \{w_h \in \mathbf{L}^2(\Omega) : w_h|_\kappa \in \mathscr{P}^p(\kappa) \quad \forall \kappa \in \Omega\}$. Note that traces of functions $u_h \in W_h^p$ are in general double valued on each edge, $e \in \mathscr{E}^i$, with values u_h^+ and u_h^- corresponding to traces from elements κ^+ and κ^- respectively. On $e \in \mathscr{E}^\partial$, associate u_h^+ with the trace taken from the element, $\kappa^+ \in \mathscr{T}_h$, neighbouring e. The weak form of (1) on each element is given by: $\forall w_h \in \mathscr{P}^p(\kappa)$

$$(\rho \nabla u_h, \nabla w_h)_\kappa - \left\langle \rho (u_h^+ - \hat{u}_h)\boldsymbol{n}^+, \nabla w_h^+ \right\rangle_{\partial\kappa} + \left\langle \hat{\boldsymbol{q}}_h, w_h^+ \boldsymbol{n}^+ \right\rangle_{\partial\kappa} = (f, w_h)_\kappa \tag{2}$$

where $(\cdot,\cdot)_\kappa := \int_\kappa$ and $\langle \cdot,\cdot \rangle_{\partial\kappa} := \int_{\partial\kappa}$. Superscript $^+$ is used to explicitly denote values on $\partial\kappa$, taken from κ. For all $w_h \in W_h^p$, $\hat{w}_h = \hat{w}_h(w_h^+, w_h^-)$ is a single valued numerical trace on $e \in \mathscr{E}^i$, while $\hat{w}_h = 0$ for $e \in \mathscr{E}^\partial$. Note that $\hat{u}_h = 0$ on $e \in \mathscr{E}^\partial$, corresponds to weakly enforced homogeneous boundary conditions on $\partial\Omega$. Similarly $\hat{\boldsymbol{q}} = \hat{\boldsymbol{q}}(\rho^+, \rho^-, \nabla u_h^+, \nabla u_h^-, u_h^+, u_h^-)$ is a single valued numerical flux approximating $\boldsymbol{q} = \rho \nabla u$ on $e \in \mathscr{E}$. Summing over all elements gives:

$$a(u_h, w_h) = (f, w_h)_\Omega \qquad\qquad \forall w_h \in W_h^p \tag{3}$$

A key component, required for the development and analysis of the algorithms presented, is to express the global bilinear form $a(u_h, w_h)$ as the sum of element-wise contributions $a_\kappa(u_h, w_h)$ such that

$$a(u_h, w_h) = \sum_{\kappa \in \mathscr{T}} a_\kappa(u_h, w_h) \tag{4}$$

where $a_\kappa(u_h, w_h)$ is a symmetric, positive semi-definite "local bilinear form". In particular, the local bilinear form should have a compact stencil, such that $a_\kappa(u_h, w_h)$ is a function of only u_h, ∇u_h in κ, and u_h^+, ∇u_h^+ and $\hat{u}_h$ on $\partial\kappa$. The local bilinear form is written as:

$$\begin{aligned} a_\kappa(u_h, w_h) &= (\rho \nabla u_h, \nabla w_h)_\kappa - \left\langle \rho (u_h^+ - \hat{u}_h)\boldsymbol{n}^+, \nabla w_h^+ \right\rangle_{\partial\kappa} + \left\langle \hat{\boldsymbol{q}}_h^+, (w_h^+ - \hat{w}_h)\boldsymbol{n}^+ \right\rangle_{\partial\kappa} \\ &= (\rho \nabla u_h, \nabla w_h)_\kappa - \left\langle \rho \left[\!\left[u \right]\!\right]_h^+, \nabla w_h^+ \right\rangle_{\partial\kappa} + \left\langle \hat{\boldsymbol{q}}_h^+, \left[\!\left[w_h \right]\!\right]^+ \right\rangle_{\partial\kappa} \end{aligned} \tag{5}$$

where $\hat{\boldsymbol{q}}_h^+ = \hat{\boldsymbol{q}}_h^+(\rho^+, \nabla u_h^+, u_h^+, \hat{u}_h)$ is a "local numerical flux". The choice of the numerical trace $\hat{u}_h$ and flux $\hat{\boldsymbol{q}}_h$ define the particular DG method considered. Table 1 lists the numerical traces and fluxes for the DG methods considered in this paper, while Table 2 lists the corresponding local bilinear forms.

DG Method	$\hat{u}_h$	$\hat{\boldsymbol{q}}_h$	$\hat{\boldsymbol{q}}_h^+$
IP	$\{u_h\}$	$-\{\rho\nabla u_h\} + \frac{\eta_e}{h}\{\rho [\![u_h]\!]^{\pm}\}$	$-\rho^+\nabla u_h^+ + \frac{\eta_e}{h}\rho^+ [\![\rho u_h]\!]^+$
BR2	$\{u_h\}$	$-\{\rho\nabla u_h\} + \eta_e\{\rho r_e([\![u_h]\!]^{\pm})\}$	$-\rho^+\nabla u_h^+ + \eta_e\rho^+ r_e([\![u_h]\!]^+)$
Brezzi	$\{u_h\}$	$\{\boldsymbol{q}_h\} + \eta_e\{\rho r_e([\![u_h]\!]^{\pm})\}$	$\boldsymbol{q}_h^+ + \eta_e\rho^+ r_e([\![u_h]\!]^+)$
LDG	$\{u_h\} - \beta\cdot[\![u_h]\!]$	$\{\boldsymbol{q}_h\} + \beta [\![\boldsymbol{q}_h]\!] + \frac{2\eta_e}{h}\{\rho [\![u_h]\!]^{\pm}\}$	$\boldsymbol{q}_h^+ + \frac{\eta_e}{h}\rho^+ [\![u_h]\!]^+$
CDG	$\{u_h\} - \beta\cdot[\![u_h]\!]$	$\{\boldsymbol{q}_h^e\} + \beta [\![\boldsymbol{q}_h^e]\!] + \frac{2\eta_e}{h}\{\rho [\![u_h]\!]^{\pm}\}$	$\boldsymbol{q}_h^{e+} + \frac{\eta_e}{h}\rho^+ [\![u_h]\!]^+$

Table 1. Numerical fluxes for different DG methods. (IP: Interior Penalty, BR2: [3], Brezzi: [4], LDG: [5] CDG: [16])

Method $a_\kappa(u_h, w_h)$	
IP	$g + \sum_{e\in\partial\kappa}\frac{\eta_e}{h_e}\left\langle \rho [\![u_h]\!]^+, [\![w_h]\!]^+\right\rangle_e$
BR2	$g + \sum_{e\in\partial\kappa}\eta_e\left(\rho r_e([\![u_h]\!]^+), r_e([\![w_h]\!]^+)\right)_\kappa$
Brezzi	$g + \left(\rho r_\kappa([\![u_h]\!]^+), r_\kappa([\![w_h]\!]^+)\right)_\kappa + \sum_{e\in\partial\kappa}\eta_e\left(\rho r_e([\![u_h]\!]^+), r_e([\![w_h]\!]^+)\right)_\kappa$
LDG	$g + \left(\rho r_\kappa([\![u_h]\!]^+), r_\kappa([\![w_h]\!]^+)\right)_\kappa + \sum_{e\in\partial\kappa}\frac{\eta_e}{h_e}\left\langle \rho [\![u_h]\!]^+, [\![w_h]\!]^+\right\rangle_e$
CDG	$g + \sum_{e\in\partial\kappa}\left(\rho r_e([\![u_h]\!]^+), r_e([\![w_h]\!]^+)\right)_\kappa + \sum_{e\in\partial\kappa}\frac{\eta_e}{h_e}\left\langle \rho [\![u_h]\!]^+, [\![w_h]\!]^+\right\rangle_e$
Where $g = (\rho\nabla u_h, \nabla w_h)_\kappa - \left\langle \rho [\![u_h]\!]^+, \nabla w_h^+\right\rangle_{\partial\kappa} - \left\langle \rho\nabla u_h, [\![w_h]\!]^+\right\rangle_{\partial\kappa}$	

Table 2. Elementwise bilinear form for different DG methods

In the definition of the different DG methods, $\{u_h\} = \frac{1}{2}(u_h^+ + u_h^-)$ and $[\![u_h]\!] = u_h^+\boldsymbol{n}^+ + u_h^-\boldsymbol{n}^-$ are average and jump operators on $e \in \mathscr{E}^i$. Additionally, a second set of jump operators involving the numerical trace $\hat{u}$ are given by $[\![u_h]\!]^+ = u_h^+\boldsymbol{n}^+ + \hat{u}_h\boldsymbol{n}^-$ and $[\![u_h]\!]^- = \hat{u}_h\boldsymbol{n}^+ + u_h^-\boldsymbol{n}^-$. Define $\boldsymbol{q}_h = -\rho(\nabla u_h - r_\kappa([\![u_h]\!]^+))$ and $\boldsymbol{q}_h^e = -\rho(\nabla u_h - r_e([\![u]\!]^+))$ where $r_\kappa(\phi)$ and $r_e(\phi) \in [\mathscr{P}^p(\kappa)]^n$ are lifting operators defined such that: $(r_\kappa(\phi), \boldsymbol{v}_h)_\kappa = \left\langle \phi, \boldsymbol{v}_h^+\right\rangle_\kappa$ and $(r_e(\phi), \boldsymbol{v}_h)_\kappa = \left\langle \phi, \boldsymbol{v}_h^+\right\rangle_e$, $\forall \boldsymbol{v}_h \in [\mathscr{P}^p(\kappa)]^n$. Additionally, on each edge in $\mathscr{E}$, η_e is a penalty parameter, while $\beta = \frac{1}{2}S_{\kappa^+}^{\kappa^-}\boldsymbol{n}^+ + S_{\kappa^-}^{\kappa^+}\boldsymbol{n}^-$ is a vector where $S_{\kappa^+}^{\kappa^-} \in \{0,1\}$ is a switch defined, such that $S_{\kappa^+}^{\kappa^-} + S_{\kappa^-}^{\kappa^+} = 1$.

Consider using a nodal basis on each element κ to define W_h^p. Figure 1 shows graphically the nodal degrees of freedom involved in defining the local bilinear form. For the IP, BR2 and Brezzi schemes, the numerical trace $\hat{u}_h$ on an edge/face depends on both u_h^+ and u_h^-. Hence the local bilinear form corresponds to all nodal degrees of freedom defining u_h on κ as well as nodal values on all edge/faces of $\partial\kappa \cap \mathscr{E}^i$

corresponding to the trace of u_h from elements neighbouring κ. On the other hand, for the LDG and CDG methods, the numerical trace $\hat{u}_h$ takes on the value of u_h^+ if $S^{\kappa^-}_{\kappa^+} = 0$ or u_h^- if $S^{\kappa^-}_{\kappa^+} = 1$. Hence the local bilinear form corresponds only to degrees of freedom defining u_h on κ and nodal values corresponding to the trace of u_h on neighbouring elements across edge/faces of $\partial\kappa \cap \mathscr{E}^i$ for which $S^{\kappa^-}_{\kappa^+} = 1$.

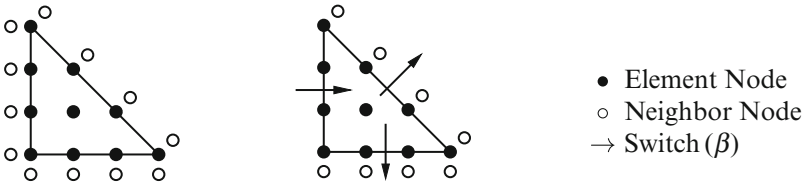


Fig. 1. Degrees of freedom involved in "local" bilinear form

The element-wise bilinear form $a_\kappa(u_h, u_h)$ satisfies

$$a_\kappa(u_h, u_h) \geq 0 \tag{6}$$

with $a_\kappa(u_h, u_h) = 0$ iff $u_h = \hat{u}_h = K$ for some constant K. The proof of (6) closely follows the proof of boundedness and stability of the different DG methods presented in [2]. As a result it is possible to show that the bilinear form is equivalent to a quadratic form based on the value of u_h at the nodes $\boldsymbol{x}$:

$$ca_\kappa(u_h, u_h) \leq \rho_\kappa p^4 h^{n-2} \sum_{\boldsymbol{x}_i, \boldsymbol{x}_j \in \kappa \cup \kappa'} (u_h(\boldsymbol{x}_i) - u_h(\boldsymbol{x}_j))^2 \leq Ca_\kappa(u_h, u_h) \tag{7}$$

where c and C are constants independent of h, p and ρ, while $\boldsymbol{x}_i, \boldsymbol{x}_j$ are the nodes on κ defining the basis for u_h and nodes on $\partial\kappa'$ defining a basis for the trace u_h^- from neighbours κ' of κ. Using the quadratic form in (7) a connection may be drawn between the DG discretization a continuous finite element discretization on a subtriangulation (See for example [6] Lemma 4.3). Further details are given in [7].

3 Domain Decomposition

Consider a partition of the domain Ω into substructures Ω_i such that $\bar{\Omega} = \cup_{i=1}^N \bar{\Omega}_i$. The substructures Ω_i are disjoint shape regular polygonal regions of diameter $O(H)$, consisting of a union of elements in $\mathscr{T}$. Assume that $\rho(\boldsymbol{x})$ takes on a constant value, ρ_i, within each subdomain Ω_i. Additionally, assume that each element κ in Ω_i with an edge e on $\partial\Omega_i \cap \partial\Omega_j$ has neighbours only in $\Omega_i \cup \Omega_j$.

Define the local interface $\Gamma_i = \partial\Omega_i \backslash \partial\Omega$ and global interface Γ by $\Gamma = \cup_{i=1}^N \Gamma_i$. Denote by $W_\Gamma^{(i)}$ the space of discrete nodal values on Γ_i which correspond to degrees

of freedom shared between Ω_i and neighbouring subdomains Ω_j, while $W_I^{(i)}$ denotes the space of discrete unknowns local to a single substructure Ω_i. In particular, note that for the IP, BR2 and Brezzi et al. methods $W_\Gamma^{(i)}$ includes for each edge $e \in \Gamma_i$ degrees of freedom defining two sets of trace values u^+ from $\kappa^+ \in \Omega_i$ and u^- for $\kappa^- \in \Omega_j$. Thus, $W_I^{(i)}$ corresponds to nodal values strictly interior to Ω_i or on $\partial\Omega_i \backslash \Gamma_i$. On the other hand, for the CDG and LDG methods $W_\Gamma^{(i)}$ includes for each edge $e \in \Gamma_i$ degrees of freedom defining a single trace value corresponding to either u^+ from $\kappa^+ \in \Omega_i$ if $S_{\kappa^+}^{\kappa^-} = 0$ or u^- from $\kappa^- \in \Omega_j$ if $S_{\kappa^+}^{\kappa^-} = 1$. Hence, $W_I^{(i)}$ corresponds to nodal values interior to Ω_i and on $\partial\Omega_i \backslash \Gamma_i$ as well as nodal values defining u^+ on $e \in \Gamma_i$ for which $S_{\kappa^+}^{\kappa^-} = 1$.

Similarly, define $\hat{W}_\Gamma$ as the space of degrees of freedom shared among multiple subdomains and W_I as the space of degrees of freedom which correspond only to a single subdomain. Note that W_I is equal to the product space $W_I := \Pi_{i=1}^N W_I^{(i)}$, while in general $\hat{W}_\Gamma \subset W_\Gamma := \Pi_{i=1}^N W_\Gamma^{(i)}$. Define local operators $R_\Gamma^{(i)} : \hat{W}_\Gamma \to W_\Gamma^{(i)}$ which extract the local degrees of freedom on Γ_i from those on Γ. Additionally define a global operator $R_\Gamma : \hat{W}_\Gamma \to W_\Gamma$ which is formed by a direct assembly of $R_\Gamma^{(i)}$. The discrete form of (3) is written as:

$$\begin{bmatrix} A_{II} & A_{\Gamma I}^T \\ A_{\Gamma I} & A_{\Gamma\Gamma} \end{bmatrix} \begin{bmatrix} u_I \\ u_\Gamma \end{bmatrix} = \begin{bmatrix} b_I \\ b_\Gamma \end{bmatrix}. \tag{8}$$

where u_I and u_Γ corresponds to degrees of freedom associated with W_I and $\hat{W}_\Gamma$ respectively. Since the degrees of freedom associated with W_I are local to a particular substructure they may be locally eliminated to obtain a system

$$\hat{S}_\Gamma u_\Gamma = g_\Gamma \tag{9}$$

where $\hat{S}_\Gamma = A_{\Gamma\Gamma} - A_{\Gamma I} A_{II}^{-1} A_{\Gamma I}^T$ and $g_\Gamma = b_{\Gamma\Gamma} - A_{\Gamma I} A_{II}^{-1} b_{\Gamma I}$. $\hat{S}_\Gamma$ and g_Γ may be formed by a direct assembly:

$$\hat{S}_\Gamma = \sum_{i=1}^N R_\Gamma^{(i)^T} S_\Gamma^{(i)} R_\Gamma^{(i)} \qquad g_\Gamma = \sum_{i=1}^N R_\Gamma^{(i)^T} g_\Gamma^{(i)} \tag{10}$$

where $S_\Gamma^{(i)} = A_{\Gamma\Gamma}^{(i)} - A_{\Gamma I}^{(i)} A_{II}^{(i)^{-1}} A_{\Gamma I}^{(i)^T}$ and $g_\Gamma^{(i)} = b_\Gamma^{(i)} - A_{\Gamma I}^{(i)} A_{II}^{(i)^{-1}} b_I^{(i)}$.

4 BDDC

A BDDC preconditioner is used to solve the Schur complement problem (9). A full description of the BDDC preconditioner is given by Li and Widlund [12]. In order to define the BDDC preconditioner $W_\Gamma^{(i)}$ is reparameterize into two orthogonal spaces $W_\Pi^{(i)}$ and $W_\Delta^{(i)}$. The primal space $W_\Pi^{(i)}$ is the space of discrete unknowns corresponding to functions with a constant value of $\hat{u}$ on each edge of substructure Ω_i. The

dual space, $W_\Delta^{(i)}$ is the space of discrete unknowns corresponding to functions which have zero mean value of $\hat{u}$ on Γ_i. For continuous finite element discretizations, different primal degrees of freedom such as subdomain corners have also been used, however these are not explored in this work. The BDDC algorithm is implemented using a change of basis as described in [12]. The partially assembled space is defined as $\tilde{W}_\Gamma = \hat{W}_\Pi \oplus \left(\Pi_{i=1}^N W_\Delta^{(i)}\right)$, where $\hat{W}_\Pi$, single valued on Γ, is formed by assembling the local primal spaces, $W_\Pi^{(i)}$. Define additional local operators $\bar{R}_\Gamma^{(i)} : \tilde{W}_\Gamma \to W_\Gamma^{(i)}$ which extract the degrees of freedom in $\tilde{W}_\Gamma$ corresponding to Γ_i. The global operator $\bar{R}_\Gamma : \tilde{W}_\Gamma \to W_\Gamma$ is formed by a direct assembly of $\bar{R}_\Gamma^{(i)}$. Also define the global operator $\tilde{R}_\Gamma : \hat{W}_\Gamma \to \tilde{W}_\Gamma$. The partially assembled Schur complement matrix $\tilde{S}$, is given by:

$$\tilde{S}_\Gamma = \sum_{i=1}^N \bar{R}_\Gamma^{(i)^T} S_\Gamma^{(i)} \bar{R}_\Gamma^{(i)} \tag{11}$$

The scaled operator $\tilde{R}_{D,\Gamma} : \hat{W}_\Gamma \to \tilde{W}_\Gamma$ is obtained by multiplying the entries of $\tilde{R}_\Gamma$ corresponding to $W_\Delta^{(i)}$ by $\delta_i^\dagger(x)$, where $\delta_i^\dagger(x)$ defined for each nodal degree of freedom in $W_\Gamma^{(i)}$ on $\partial\Omega_i$ and $\partial\Omega_j$ as $\delta_i^\dagger = \frac{\rho_i^\gamma}{\rho_i^\gamma + \rho_j^\gamma}$, $\gamma \in [1/2, \infty)$. The BDDC preconditioner $M_{\text{BDDC}}^{-1} : \hat{W}_\Gamma \to \hat{W}_\Gamma$ is given by:

$$M_{\text{BDDC}}^{-1} = \tilde{R}_{D,\Gamma}^T \tilde{S}_\Gamma^{-1} \tilde{R}_{D,\Gamma} \tag{12}$$

The condition number of the preconditioner operator $M_{\text{BDDC}}^{-1}\hat{S}$ is bounded by $C(1 + \log(p^2 H/h))^2$ where C is a constant independent of p, h, H or ρ. This is the same condition number bound as obtained by Klawonn et al. [11] for a continuous finite element discretization. Proof of this condition number bound closely follows that presented by Tu [17] for mixed finite element methods, which in turn builds upon the work of [6]. The key idea is to connect the DG discretization to a related continuous finite element discretization on a subtriangulation of $\mathscr{T}$. The ability to connect the DG discretization to the continuous finite element discretization is a direct result of (7) (see [6]). The existing theory for continuous finite elements developed in [13, 15] and [11] is then leveraged to obtain the desired condition number bound. Further details are provided in [7].

5 Numerical Results

This section presents numerical results using the BDDC preconditioner introduced in Sect. 4. For each numerical experiment the linear system resulting from the DG discretization is solved iteratively using a Preconditioned Conjugate Gradient (PCG) method, starting from zero initial condition until l_2 norm of the residual is decreased by a factor of 10^{10}. The domain $\Omega = (0,1)^2$ is partitioned into $N \times N$ square subdomains Ω_i with side lengths H such that $N = \frac{1}{H}$. Each subdomain is the union of triangular elements obtained by bisecting squares of side length h. In the first numerical

experiment (1) is solved on Ω with $\rho = 1$ and f chosen such that the exact solution is given by $u = \sin(\pi x)\sin(\pi y)$. Table 3 shows the number of PCG iteration required to converge varying N, $\frac{H}{h}$ and p for each of the DG discretization considered. Table 3 also gives the Lanczos estimate of the maximum eigenvalue of the preconditioned system. The minimum eigenvalue is bounded below by unity as with continuous finite element methods. As expected the number of iterations is independent of the number of subdomains and only weakly dependent on the number of elements per subdomain or the solution order.

$\frac{1}{H}$	$\frac{H}{h}$	p	IP	BR2	Brezzi	LDG	CDG
2			12 (12.1)	15 (12.0)	15 (7.7)	11 (6.1)	12 (5.9)
4			22 (14.3)	27 (14.0)	23 (9.2)	24 (7.4)	24 (7.1)
8	8	4	31 (15.2)	34 (14.8)	30 (9.8)	28 (7.7)	27 (7.5)
16			33 (15.3)	36 (14.9)	32 (9.9)	29 (8.0)	28 (7.8)
32			33 (15.3)	36 (14.9)	32 (9.9)	29 (7.9)	27 (7.7)
	2		25 (10.9)	29 (10.9)	26 (6.9)	23 (5.2)	23 (5.3)
	4		29 (13.0)	34 (12.8)	28 (8.3)	26 (6.4)	25 (6.2)
8	8	4	31 (15.2)	34 (14.8)	30 (9.8)	28 (7.8)	27 (7.5)
	16		33 (17.6)	36 (17.1)	33 (11.5)	29 (9.3)	29 (9.1)
	32		35 (20.2)	38 (19.4)	34 (13.4)	32 (11.0)	31(10.7)
		1	32 (11.1)	36 (13.8)	28 (8.1)	26 (5.9)	25 (5.6)
		2	31 (12.9)	34 (14.1)	29 (8.7)	26 (6.4)	26 (6.3)
8	8	4	31 (15.2)	34 (14.8)	30 (9.8)	28 (7.8)	27 (7.5)
		8	34 (18.4)	37 (16.2)	34 (11.7)	31 (9.9)	32 (9.6)
		16	36 (22.5)	38 (18.6)	38 (14.4)	34 (12.8)	36 (12.2)

Table 3. Iteration count ($\lambda_{\max}$) for BDDC preconditioner using different DG methods

In the second numerical experiment the behaviour of the preconditioner for large jumps in the coefficient ρ is examined. For this numerical experiment only the CDG discretization is used. The domain is partitioned in a checkerboard pattern with $\rho = 1$ on half of the subdomains and $\rho = 1{,}000$ in the remaining subdomains. Initially set $\delta_i^\dagger = \frac{1}{2}$, which corresponds to setting $\gamma = 0$, which does not satisfy the assumption $\gamma \in [1/2, \infty)$. Poor convergence of the BDDC algorithm is seen in Table 4a. Next $\delta_i^\dagger$ is set to $\delta_i^\dagger = \frac{\rho_i}{\rho_i + \rho_j}$ which corresponds to $\gamma = 1$. With this choice of $\delta_i^\dagger$ the good convergence properties of the BDDC algorithm is recovered as shown in Table 4b.

6 Conclusions

The BDDC preconditioner has been extended to a large class of DG discretizations for second-order elliptic problems. The condition number of the BDDC preconditioned system is bounded by $C(1 + \log(p^2 H/h))^2$, with constant C independent of p, h, H or the coefficient ρ. This is the same condition number bound previously proven for continuous finite element methods. Numerical results confirm the theory.

(a) $\delta_i^\dagger = \frac{1}{2}$, $\frac{H}{h} = 8$

	$\frac{1}{H}$				
p	2	4	8	16	32
1	51	119	179	215	232
3	55	133	207	267	316
5	59	153	242	306	361

(b) $\delta_i^\dagger = \frac{\rho_i}{\rho_i+\rho_j}$, $\frac{H}{h} = 8$

	$\frac{1}{H}$				
p	2	4	8	16	32
1	4	7	14	18	19
3	4	7	15	18	19
5	4	7	14	19	20

Table 4. Iteration count for BDDC preconditioner using the CDG method with $\rho = 1$ or 1000.

Bibliography

[1] P. Antonietti and B. Ayuso. Schwarz domain decomposition preconditioners for discontinuous Galerkin approximations of elliptic problems: Non-overlapping case. *Math. Model. Numer. Anal.*, 41(1):21–54, 2007.

[2] D. Arnold, F. Brezzi, B. Cockburn, and D. Marini. Unified analysis of discontinuous Galerkin methods for elliptic problems. *SIAM J. Numer. Anal.*, 39(5): 1749–1779, 2002.

[3] F. Bassi and S. Rebay. A high-order discontinuous finite element method for the numerical solution of the compressible Navier-Stokes equations. *J. Comput. Phys.*, 131:267–279, 1997.

[4] F. Brezzi, G. Manzini, D. Marini, P. Pietra, and A. Russo. Discontinuous Galerkin approximations for elliptic problems. *Numer. Meth. Part. D. E.*, 16 (4):365–378, July 2000.

[5] B. Cockburn and C.W. Shu. The local discontinuous Galerkin method for time-dependent convection-diffusion systems. *SIAM J. Numer. Anal.*, 35(6): 2440–2463, December 1998.

[6] L. Cowsar, J. Mandel, and M. Wheeler. Balancing domain decomposition for mixed finite elements. *Math. Comp.*, 64(211):989–1015, 1995.

[7] Laslo T. Diosady and David L. Darmofal. A Unified Analysis of Balancing Domain Decomposition by Constraints for Discontinuous Galerkin Discretizations. *SIAM J. Numer. Anal.*, 50(3):1695–1712, 2012.

[8] Clark R. Dohrmann. A preconditioner for substructuring based on constrained energy minimization. *SIAM J. Sci. Comput.*, 25(1):246–258, 2003.

[9] M. Dryja, J. Galvis, and M. Sarkis. BDDC methods for discontinuous Galerkin discretization of elliptic problems. *J. Complexity*, 23(4):715–739, 2007.

[10] X. Feng and O.A. Karakashian. Two-level additive Schwarz methods for a discontinuous Galerkin approximation of second order elliptic problems. *SIAM J. Numer. Anal.*, 39(4):1343–1365, 2002.

[11] Axel Klawonn, Luca F. Pavarino, and Oliver Rheinbach. Spectral element FETI-DP and BDDC preconditioners with multi-element subdomains. *Comput. Methods Appl. Mech. Engrg.*, 198:511–523, 2008.

[12] J. Li and O.B. Widlund. FETI-DP, BDDC, and block Cholesky methods. *Internat. J. Numer. Methods Engrg.*, 66:250–271, 2006.

[13] Jan Mandel and Clark R. Dohrmann. Convergence of a balancing domain decomposition by constraints and energy minimization. *Numer. Linear Algebra Appl.*, 10:639–659, 2003.

[14] Jan Mandel and B. Sousedik. BDDC and FETI-DP under minimalist assumptions. *Computing*, 81:269–280, 2007.

[15] L.F. Pavarino. BDDC and FETI-DP preconditioners for spectral element discretizations. *Comput. Methods Appl. Mech. Engrg.*, 196(8):1380–1388, 2007.

[16] J. Peraire and P-O. Persson. The compact discontinuous Galerkin (CDG) method for elliptic problems. *SIAM J. Sci. Comput.*, 30(4):1806–1824, 2008.

[17] X. Tu. A BDDC algorithm for flow in porous media with a hybrid finite element discretization. *Electron. Trans. Numer. Anal.*, 26:146–160, 2007.

ARAS2 Preconditioning Technique for CFD Industrial Cases

Thomas Dufaud[1] and Damien Tromeur-Dervout[1]

[1] Université de Lyon - Université Lyon 1 - CNRS - UMR 5208 - Institut Camille Jordan, 43 Bd du 11 Novembre 1918, F-69622 Villeurbanne Cedex
thomas.dufaud@univ-lyon1.fr
[2] damien.tromeur-dervout@univ-lyon1.fr

1 Introduction

The convergence rate of a Krylov method such as the Generalized Conjugate Residual (GCR) [6] method, to solve a linear system $Au = f, A = (a_{ij}) \in \mathbb{R}^{m\times m}, u \in \mathbb{R}^m, f \in \mathbb{R}^m$, decreases with increasing condition number $\kappa_2(A) = ||A||_2||A^{-1}||_2$ of the non singular matrix A. Left preconditioning techniques consist of solving $M^{-1}Au = M^{-1}f$ such that $\kappa_2(M^{-1}A) << \kappa_2(A)$. The Additive Schwarz (AS) preconditioning is built from the adjacency graph $G = (W,E)$ of A, where $W = \{1,2,\ldots,m\}$ and $E = \{(i,j) : a_{ij} \neq 0\}$ are the edges and vertices of G. Starting with a non-overlapping partition $W = \cup_{i=1}^{p} W_{i,0}$ and $\delta \geq 0$ given, the overlapping partition $\{W_{i,\delta}\}$ is obtained defining p partitions $W_{i,\delta} \supset W_{i,\delta-1}$ by including all the immediate neighboring vertices of the vertices in the partition $W_{i,\delta-1}$. Then the restriction operator $R_{i,\delta}$ from W to $W_{i,\delta}$ defines the local operator $A_{i,\delta} = R_{i,\delta}AR_{i,\delta}^T, A_{i,\delta} \in \mathbb{R}^{m_{i,\delta}\times m_{i,\delta}}$ on $W_{i,\delta}$. The AS preconditioning writes: $M_{AS,\delta}^{-1} = \sum_{i=1}^{p} R_{i,\delta}^T A_{i,\delta}^{-1} R_{i,\delta}$. Introducing $\tilde{R}_{i,\delta}$ the restriction matrix on a non-overlapping subdomain $W_{i,0}$, the Restricted Additive Schwarz (RAS) iterative process [2] writes:

$$u^k = u^{k-1} + M_{RAS,\delta}^{-1}\left(f - Au^{k-1}\right), \text{with} M_{RAS,\delta}^{-1} = \sum_{i=1}^{p} \tilde{R}_{i,\delta}^T A_{i,\delta}^{-1} R_{i,\delta} \quad (1)$$

The RAS exhibits a faster convergence than the AS, as shown in [5], leading to a better preconditioning that depends of the number of subdomains. When it is applied to linear problems, the RAS has a pure linear rate of convergence/divergence that can be enhanced with optimized boundary conditions giving the ORAS method of [11]. The RAS method's linear convergence allows its acceleration of the convergence by the Aitken's process as done in [8] for the Schwarz method.

In [4] the present authors designed the ARAS2 preconditioning technique based on the Aitken's acceleration of the convergence technique. This paper presents an approach to solve linear systems coming from CFD industrial cases. The choice of an

R. Bank et al. (eds.), *Domain Decomposition Methods in Science and Engineering XX*, 569
Lecture Notes in Computational Science and Engineering 91,
DOI 10.1007/978-3-642-35275-1_67, © Springer-Verlag Berlin Heidelberg 2013

approximation space based on the Singular Value Decomposition of the interface's solutions of the RAS iterative process presented in [14] is done. This provides a preconditioning technique that depends on the Right Hand Side but with a very low computational time and totally algebraic.

2 The ARAS2 Preconditioning Method

In what follows, we write the Aitken Restricted Additive Schwarz (ARAS) iterative process and the associated preconditioner. This preconditioner belongs to the family of the two-level preconditioner techniques (see [10, 13] and references) but the coarse grid operator uses only parts of the artificial interfaces contrary to the patch substructuring method of [7]. In this way, it can be seen as similar as the SchurRAS method of [9] but it differs because the discrete Steklov-Poincaré operator connects the coarse artificial interfaces of all the subdomains.

2.1 The ARAS and ARAS2 Preconditioner's Formulation

Let $\Gamma_i = W_{i,\delta+1} \setminus W_{i,\delta}$ be the interface associated to $W_{i,\delta}$ and $\Gamma = \cup_{i=1}^{p} \Gamma_i$ be the global interface. Then $u_{|\Gamma} \in \mathbb{R}^n$ is the restriction of the solution $u \in \mathbb{R}^m$ on the Γ interface and $e^k_{|\Gamma} = u^k_{|\Gamma} - u^\infty_{|\Gamma}$ is the error of (1) at the interface Γ. Taking into account that there exists a matrix $P \in \mathbb{R}^{n\times n}$ independent of the iterate k such that $e^k_{|\Gamma} = Pe^{k-1}_{|\Gamma}$, we can apply the Aitken's acceleration of the convergence process [8] (if $||P|| < 1$ to ensure existence of $(I_n - P)^{-1}$ for example) as follows:

$$u^\infty_{|\Gamma} = (I_n - P)^{-1} \left(u^k_{|\Gamma} - Pu^{k-1}_{|\Gamma} \right). \tag{2}$$

P can be computed analytically or numerically for a separable operator on separable geometry [8] or numerically approximated in other cases [14]. Using this property on the RAS method, we would like to write a preconditioner which includes the Aitken's acceleration process. We introduce a restriction operator $R_\Gamma \in \mathbb{R}^{n\times m}$ from W to the global artificial interface Γ, with $R_\Gamma R_\Gamma^T = I_n$.

The Aitken Restricted Additive Schwarz (ARAS) must generate a sequence of solutions on the interface Γ, and accelerate the convergence of the Schwarz process from this original sequence. Then the accelerated solution on the interface replaces the last one. This could be written combining an AS or RAS process Eq. (3a) with the Aitken process written in $\mathbb{R}^{m\times m}$ Eq. (3b) and substracting the Schwarz solution which is not extrapolated on Γ Eq. (3c). We can write the following approximation u^* of the solution u:

$$u^* = u^{k-1} + M^{-1}_{RAS,\delta}(f - Au^{k-1}) \tag{3a}$$

$$+ R_\Gamma^T (I_n - P)^{-1} \left(u^k_{|\Gamma} - Pu^{k-1}_{|\Gamma} \right) \tag{3b}$$

$$- R_\Gamma^T I_n R_\Gamma \left(u^{k-1} + M^{-1}_{RAS,\delta}(f - Au^{k-1}) \right) \tag{3c}$$

We would like to write u^* as an iterated solution derived from an iterative process of the form $u^* = u^{k-1} + M_{ARAS,\delta}^{-1}\left(f - Au^{k-1}\right)$, where $M_{ARAS,\delta}^{-1}$ is the Aitken-RAS preconditioner.

Hence the formulation Eq. (3) leads to an expression of an iterated solution u^*:

$$u^* = u^{k-1} + \left(I_m + R_\Gamma^T\left((I_n - P)^{-1} - I_n\right)R_\Gamma\right)M_{RAS,\delta}^{-1}\left(f - Au^{k-1}\right)$$

This iterated solution u^* can be seen as an accelerated solution of the RAS iterative process. Drawing our inspiration from the Stephensen's method, we build a new sequence of iterates from the solutions accelerated by the Aitken's acceleration method. Such a process is done in [12]. Then, one considers u^* as a new u^k and writes the following ARAS iterative process:

$$u^k = u^{k-1} + \left(I_m + R_\Gamma^T\left((I_n - P)^{-1} - I_n\right)R_\Gamma\right)M_{RAS,\delta}^{-1}\left(f - Au^{k-1}\right) \tag{4}$$

Then we defined the ARAS preconditioner as

$$M_{ARAS,\delta}^{-1} = \left(I_m + R_\Gamma^T\left((I_n - P)^{-1} - I_n\right)R_\Gamma\right)\sum_{i=1}^{p}\tilde{R}_{i,\delta}^T A_{i,\delta}^{-1}R_{i,\delta} \tag{5}$$

If P is known exactly, the ARAS process written in Eq. (4) needs two steps to converge to the solution u with an initial guess $u^0 = 0$. Then we have:

Proposition 1. *If P is known exactly then we have* $A^{-1} = \left(2M_{ARAS,\delta}^{-1} - M_{ARAS,\delta}^{-1}AM_{ARAS,\delta}^{-1}\right)$ *that leads* $\left(I - M_{ARAS,\delta}^{-1}A\right)$ *to be a nilpotent matrix of degree 2.*

The previous proposition leads to an approximation of A^{-1} written from the 2 first iterations of the ARAS iterative process (4). Those 2 iterations compute the Schwarz solutions sequence on the interface needed in order to accelerate the Schwarz method by the Aitken's acceleration. We now write 2 iterations of the ARAS iterative process (4) for any initial guess and for all $u^{k-1} \in \mathbb{R}^m$.

$$u^{k+1} = u^{k-1} + \left(2M_{ARAS,\delta}^{-1} - M_{ARAS,\delta}^{-1}AM_{ARAS,\delta}^{-1}\right)\left(f - Au^{k-1}\right)$$

Then we defined the ARAS2 preconditioner as

$$M_{ARAS2,\delta}^{-1} = 2M_{ARAS,\delta}^{-1} - M_{ARAS,\delta}^{-1}AM_{ARAS,\delta}^{-1} \tag{6}$$

Hence, if P is known exactly there is no need to use ARAS as a preconditioning technique. Nevertheless, when P is approximated, the Aitken's acceleration of the convergence depends on the local domain solving accuracy, and the cost of the building of an exact P depends on the size n. This is why P is numerically approximated by $P_{\mathbb{U}_q}$, defining $q \leq n$ orthogonal vectors $\mathbb{U}_q \in R^{n\times q}$, that are able to approximate most of the solution at the interface Γ. Then ARAS($\mathbb{U}_q$) and ARAS2($\mathbb{U}_q$) can be defined as:

$$M^{-1}_{ARAS(\mathbb{U}_q),\delta} = \left(I_m + R_\Gamma^T \mathbb{U}_q \left(\left(I_q - P_{\mathbb{U}_q}\right)^{-1} - I_q \right) \mathbb{U}_q^T R_\Gamma \right) \sum_{i=1}^{p} \tilde{R}_{i,\delta}^T A_{i,\delta}^{-1} R_{i,\delta} \tag{7}$$

and

$$M^{-1}_{ARAS2(\mathbb{U}_q),\delta} = 2M^{-1}_{ARAS(\mathbb{U}_q),\delta} - M^{-1}_{ARAS(\mathbb{U}_q),\delta} A M^{-1}_{ARAS(\mathbb{U}_q),\delta} \tag{8}$$

As the basis $\mathbb{U}_q$ can only give an approximation of the searched solution at the interface, it make sense to use $M^{-1}_{ARAS(\mathbb{U}_q),\delta}$ and $M^{-1}_{ARAS2(\mathbb{U}_q),\delta}$ as preconditioners.

2.2 Orthogonal Basis $\mathbb{U}_q$ Arising from SVD of the Interface's Solutions of Richardson Process

The objective is to compute $P_{\mathbb{U}_q}$ saving as much computing as possible. The singular value decomposition offers a tool to concentrate the effort only on the main parts of the solution. A singular-value decomposition of a real $n \times q$ $(n > q)$ matrix Y is its factorization into the product of three matrices $Y = \mathbb{U}_q \Sigma \mathbb{V}^*$, where $\mathbb{U}_q = [U_1, \ldots, U_q]$ is an $n \times q$ matrix with orthonormal columns, Σ is an $n \times q$ nonnegative diagonal matrix with $\Sigma_{ii} = \sigma_i$, $1 \leq i \leq q$ and the $q \times q$ matrix $\mathbb{V} = [V_1, \ldots, V_q]$ is orthogonal. The left $\mathbb{U}_q$ and right $\mathbb{V}$ singular vectors are the eigenvectors of YY^* and Y^*Y respectively. It readily follows that $Av_i = \sigma_i u_i$, $1 \leq i \leq q$. We are going to recall some properties of the SVD. Assume that the $\sigma_i, 1 \leq i \leq q$ are ordered in decreasing order and there exists an r such that $\sigma_r > 0$ while $\sigma_r + 1 = 0$. Then A can be decomposed in a dyadic decomposition:

$$Y = \sigma_1 U_1 V_1^* + \sigma_2 U_2 V_2^* + \ldots + \sigma_r U_r V_r^*. \tag{9}$$

This means that SVD provides a way to find optimal lower dimensional approximations of a given series of data. More precisely, it produces an orthonormal basis for representing the data series in a certain least squares optimal sense.

The orthogonal "basis" $\mathbb{U}_q$ is obtained as follows. q iterations of the Richardson process $u^k = u^{k-1} + M^{-1}_{RAS,\delta}(f - Au^{k-1})$ are performed and $R_\Gamma u^k \in \mathbb{R}^n, 1 \leq k \leq q$ belonging to the interface Γ are stored in a matrix $Y \in \mathbb{R}^{n \times q}$. Then the SVD of Y is computed to obtain the matrix $\mathbb{U}_q$ with an arithmetic cost less than the one of a local solution. It leads to efficiency and low computational cost as illustrated in [1]. Nevertheless, the preconditioner ARAS2($\mathbb{U}_q$) obtained is solution dependent.

2.3 Building of the $P_{\mathbb{U}_q}$ Matrix

The matrix $P_{\mathbb{U}_q}$ can be computed as follows keeping the $q+1$ first singular values of the SVD greater than a set tolerance, we writes:

$$\mathbf{Y}_{1:q,1:q+1} = \Sigma_{1:q,1:q} \mathbb{V}^T_{1:q,1:q+1} \tag{10}$$

$$\mathbf{E}_{1:q,1:q+1} = \mathbf{Y}_{1:q,2:q+1} - \mathbf{Y}_{1:q,1:q} \tag{11}$$

$$\text{If } \mathbf{E}_{1:q,1:q} \text{ is invertible then} \tag{12}$$

$$P_{\mathbb{U}_q} = \mathbf{E}_{1:q,2:q+1}\, \mathbf{E}^{-1}_{1:q,1:q} \tag{13}$$

The previous building requires the inversion of the matrix $\mathbf{E}_{1:q,1:q}$ which can be ill conditioned. It is why the second building of matrix $P_{\mathbb{U}_q}$ that follows is prefered. Selecting the q first singular values of the SVD greater than a set tolerance, one iteration of the RAS algorithm is applied on the q the homogeneous problems where $U^i, 1 \leq i \leq q$ is set as boundary condition on the interface Γ. The result of this RAS iterate with $M^{-1}_{RAS,\delta}$ on the boundary Γ is the column of $P_{\mathbb{U}_q}$ associated with the component U_i of the basis. Let us notice that this q computing can be made in the same time considering the q right hand sides in a matrix form.

3 Numerical Experiments on 2D and 3D Industrial Problems from Navier-Stokes Equations

In this section we focus on solving linear systems coming from industrial problems with the ARAS2 preconditioning technique. The sparse matrices correspond to the assemblage of all the elementary Jacobian matrices resulting from the partial first-order derivations with respect to the conservative fluid variables of the discrete steady (real) Reynolds-averaged Navier-Stokes equations. We note here that the Jacobian matrix is non-symmetric and is non positive definite.

Table 1 summarizes the main features of the linear systems from the two cases solved. Those cases are available in the sparse matrix collection [3]. Turbulence is considered in the 2D and 3D cases. We partition the system with PARMETIS into p subdomains. We must notice that for such problems with non-elliptic operators, the ILU factorization is hazardous. Then, the preconditioner is computed from exact factorization of local operators.

Figure 1 presents for the case PR02 the convergence behaviour of the Richardson and the GMRES preconditioned by the ARAS2 preconditioner where the $P_{\mathbb{U}_q}$ is approximated by SVD. For this matrix the RAS Richardson process diverges. If the number of singular values kept is not sufficient, the ARAS2 process diverges as well. If we used 60 iterates of RAS Richardson process then the "full" $P_{\mathbb{U}_q}$ makes the ARAS2 Richardson process converge in one iterate. Nevertheless ARAS2 works quite well in both cases as a preconditioner of the GMRES method. We must notice that here we have an effective gain to use the ARAS2 instead of RAS as Richardson process. The same behavior is also retrieved when ARAS2 is used as preconditioner.

For a 3D case the number of non-zero and the band profile increase. Then solving local problems by LU factorization begins to be expensive in terms of memory. A better approach consists of solving subproblems by an iterative method. For the case RM07, we choose to solve subproblems by a GMRES preconditioned by ILU. The idea to save computational time is to approximate the Aitken's acceleration with the basis arising from SVD and solving subproblems with less accuracy for the computing of the preconditioner. Table 2 shows the good strong numerical scalability of the ARAS2 preconditioning compare to the RAS.

case ID	*order*	*dim*	*nn*	*nnz*
PR02	161 070	2*D*	23 010	8 185 136
RM07	381 689	3*D*	54 527	37 464 962

Table 1. Main features of the linear systems with *order* the size of the matrix with real coefficients, *dim* the dimension of the problem, *nn* is the number of mesh nodes, *nnz* is the number of non-zero elements in the matrix

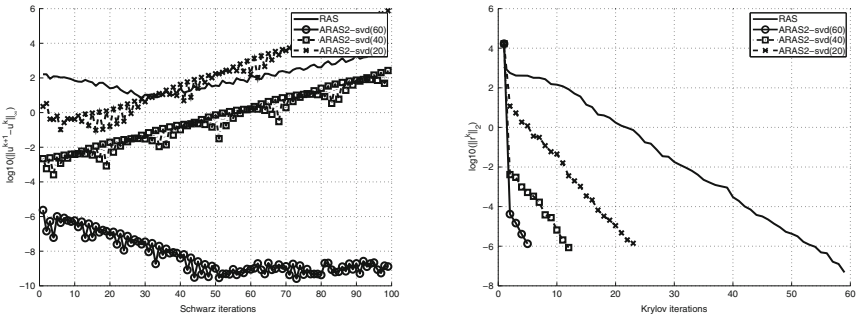


Fig. 1. Solving 2D Navier Stokes equation with turbulence (CASE PR02), PARMETIS partitioning, $p = 4$, overlap 2, ARAS2 is built with a SVD basis, (*left*) Convergence of Iterative Schwarz Process, (*right*) convergence of GMRES method preconditioned by RAS and ARAS2

p	RAS	ARAS(36)	ARAS2(36)
3	87 (1.)	77 (1.1299)	53 (1.6415)
6	112 (1.)	93 (1.2043)	63 (1.7778)
12	171 (1.)	124 (1.3790)	84 (2.0357)

Table 2. CASE RM07 : Number of GMRES iterations (ratio of iterations with RAS over iterations with ARAS or ARAS2) for a tolerance 1e-10, overlap 1.

Acknowledgments This work was funded by the French National Agency of Research under the contract ANR-TLOG07-011-03 LIBRAERO. The work of the second author was also supported by the région Rhône-Alpes through the cluster AUTOMOTIVE.

Bibliography

[1] L. Berenguer, T. Dufaud, and D. Tromeur-Dervout. Aitken's acceleration of the schwarz process using singular value decomposition for heterogeneous 3d groundwater flow problems. *Computers & Fluids*, 2012. doi: 10.1016/j.compfluid.2012.01.026. URL http://dx.doi.org/10.1016/j.compfluid.2012.01.026.

[2] X.-C. Cai and M. Sarkis. A restricted additive Schwarz preconditioner for general sparse linear systems. *SIAM J. Sci. Comput.*, 21(2):792–797 (electronic), 1999.

[3] T. A. Davis and Y. Hu. The university of florida sparse matrix collection, acm transactions on mathematical software (to appear), 2009. http://www.cise.ufl.edu/research/sparse/matrices.

[4] T. Dufaud and D. Tromeur-Dervout. Aitken's acceleration of the restricted additive Schwarz preconditioning using coarse approximations on the interface. *C. R. Math. Acad. Sci. Paris*, 348(13–14):821–824, 2010.

[5] E. Efstathiou and M. J. Gander. Why restricted additive Schwarz converges faster than additive Schwarz. *BIT*, 43(suppl.):945–959, 2003.

[6] S. C. Eisenstat, H. C. Elman, and M. H. Schultz. Variational iterative methods for nonsymmetric systems of linear equations. *SIAM J. Numer. Anal.*, 20(2): 345–357, 1983.

[7] M. J. Gander, L. Halpern, F. Magoulès, and F.-X. Roux. Analysis of patch substructuring methods. *Int. J. Appl. Math. Comput. Sci.*, 17(3):395–402, 2007.

[8] M. Garbey and D. Tromeur-Dervout. On some Aitken-like acceleration of the Schwarz method. *Internat. J. Numer. Methods Fluids*, 40(12):1493–1513, 2002. LMS Workshop on Domain Decomposition Methods in Fluid Mechanics (London, 2001).

[9] Z. Li and Y. Saad. SchurRAS: a restricted version of the overlapping Schur complement preconditioner. *SIAM J. Sci. Comput.*, 27(5):1787–1801 (electronic), 2006.

[10] A. Quarteroni and A. Valli. *Domain decomposition methods for partial differential equations*. Numerical Mathematics and Scientific Computation. The Clarendon Press Oxford University Press, New York, 1999. Oxford Science Publications.

[11] A. St-Cyr, M. J. Gander, and S. J. Thomas. Optimized multiplicative, additive, and restricted additive Schwarz preconditioning. *SIAM J. Sci. Comput.*, 29(6): 2402–2425 (electronic), 2007.

[12] J. Stoer and R. Bulirsch. *Introduction to numerical analysis*, volume 12 of *Texts in Applied Mathematics*. Springer-Verlag, New York, third edition, 2002. Translated from the German by R. Bartels, W. Gautschi and C. Witzgall.

[13] A. Toselli and O. Widlund. *Domain decomposition methods—algorithms and theory*, volume 34 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin, 2005.

[14] D. Tromeur-Dervout. Meshfree Adaptative Aitken-Schwarz Domain Decomposition with application to Darcy Flow. In Topping, BHV and Ivanyi, P, editor, *Parallel, Distributed and Grid Computing for Engineering*, volume 21 of *Computational Science Engineering and Technology Series*, pages 217–250. Saxe-Coburg Publications, 2009.

An Implicit and Parallel Chimera Type Domain Decomposition Method

B. Eguzkitza[1], G. Houzeaux[1], R. Aubry[2], and O. Peredo[1]

[1] Barcelona Supercomputing Center (BSC-CNS),
Dept. Computer Applications in Science and Engineering,
Edificio NEXUS I, Campus Nord UPC,
beatriz.eguzkitza@bsc.es

[2] CFD Center, Dept. of Computational and Data Science
M.S. 6A2, College of Science, George Mason University
Fairfax, VA 22030-4444, USA
US Naval Research Laboratory
4555 Overlook Ave SW
Washington DC 20375, USA

1 Introduction

The Chimera Method developed originally in [1, 19, 20] simplifies the construction of computational meshes about complex geometries. This is achieved by breaking the geometries into components and generating independently a series of different meshes. This enables one a great flexibility on the choice of the type of elements, their orientations and local mesh refinement. The components are further coupled by transmitting information from one mesh to the other to obtain a global solution.

The Chimera Method is a very efficient tool to treat moving objects [3, 16] as the different meshes can move as rigid bodies in an independent way. Nevertheless, we will focus in this work on fixed subdomains. The main application in this context is optimization analysis, where different configurations can be tested without having to remesh the whole geometry. In order to achieve this, we have developed a versatile strategy based on the Chimera Method.

Usually, in the Chimera Method, the mesh is divided into a background mesh, which covers all the computational domain, and patch (overset) meshes attached to the different components (objects) which are located upon the background mesh. First, we apply a proper preprocessing consisting in removing elements of the background mesh located inside the patch meshes to create apparent interfaces between the background and the patches. The present algorithm requires in addition to smooth the interfaces. This is achieved using a smoothing strategy of the interfaces and the neighboring volume mesh. Then a new coupling algorithm is carried out in order to obtain a "continuous solution" across the interfaces. In the literature, the Chimera coupling has generally been implemented as an iterative algorithm (see [2] for a

R. Bank et al. (eds.), *Domain Decomposition Methods in Science and Engineering XX*, 577
Lecture Notes in Computational Science and Engineering 91,
DOI 10.1007/978-3-642-35275-1_68,

Schwarz coupling or [9] for a Dirichlet/Neumann coupling). Here the coupling is implicit. The implementation properties of the proposed coupling facilitate its parallel implementation and makes it a versatile method to be used on general PDE's.

In the following we explain the two basic steps of the Chimera method. The preprocessing step which consists in creating the interfaces between the subdomains. This is a purely geometrical task. We then present the coupling step which couples the solution from the different meshes. Finally we show a numerical examples.

2 Interface Creation Process

The first task of the Chimera method is to create apparent interfaces between the background and the patch meshes. This is achieved by the hole cutting step of the Chimera method. As will be explained in next section, our coupling strategy requires smooth interfaces. After the hole cutting, smoothing of the interfaces are also necessary. We now explain these two points.

2.1 Hole Cutting

The hole cutting tasks consists in removing elements (the hole elements) from the background mesh to form interfaces with the patches. We start by identifying the hole nodes. The hole nodes are those nodes of the background mesh that are located inside the patch mesh. To do this we have used a *skd-tree* strategy, as explained in [12]. Skd-trees are used to find efficiently the signed shortest distance between a point and a surface. In our case, the surfaces are the patch outer boundaries. In practice we obtain a better efficiency if we use the search algorithm described in [18], which is a slightly modified version of the above reference. Having found the hole nodes, we identify the hole elements which are the background elements of which all nodes are hole nodes. The fringe nodes are defined as the nodes located on the outer boundaries of the hole elements. They are the hole nodes having non-hole neighbor nodes. The fringe nodes are used to form the interface of the backgound with the patches.

2.2 Smoothing

The domain decomposition coupling we propose is geometrical, as will be shown in next section. It is therefore important to ensure a minimum regularity of the interfaces and the mesh nearby, as this will affect the quality of the results. Figure 1 (Left) shows an example of typical background interface resulting from the previous hole cutting process. The proposed strategy consists in smoothing first the interface and then the volume mesh in the vicinity.

In this article, we are interested in mesh smoothing techniques that relocate the nodes to improve the mesh without changing its topology. The particular method we consider here is based on local mesh smoothing algorithms, since they have shown to be efficient in repairing distorted elements. The most common smoothing technique is Laplacian smoothing (see [13]), which moves a given node to the barycenter

of all its connected nodes. This method is not computationally expensive but does not guarantee an improvement in mesh quality. In addition, it can create invalid elements or poor quality elements resulting in convergence and shrinkage problems. To overcome this shortcoming, different variations of Laplacian smoothing have been proposed like [5, 22].

Optimization-based smoothing algorithms are alternative local smoothing strategies. These algorithms depend on the type of mesh, the optimization method used and a measure of the mesh quality, and require an objective function to be optimized. The objective function should include a good representation of the mesh quality. A good summary of measures for the quality of tetrahedra and a global definition of the tetrahedron shape measure is given in [4]. Besides the geometrical objective functions described in the above reference, there exist other quality interpretations based on matrices and matrix norms. This matrix perspective suggests several different objective functions as, for example, the smoothness objective function in terms of the condition number of the Jacobian matrix; see [6].

Our smoothing process consists first of a surface Laplacian-smoothing algorithm based on [21] for the interface. An example is shown in Fig. 1. As a consequence,

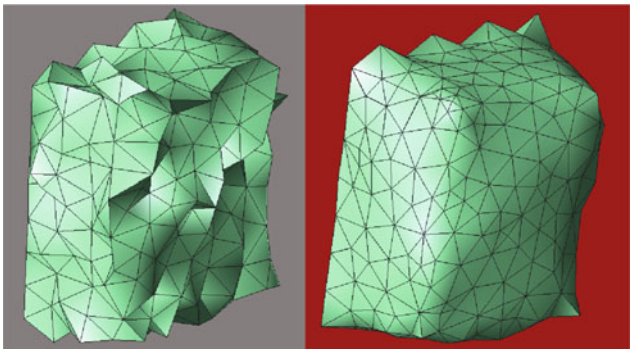


Fig. 1. (*Left*) Original interface after hole cutting. (*Right*) Smoothed interface

we need to relocate the volume nodes in order to repair the bad quality elements. To tackle this problem, we have applied a tetrahedra mesh improvement via optimization of the element condition number developed in [6]. This optimization uses a steepest descent method with a modified line search adapted to the geometrical constraints of the sub-mesh associated to the node we want to move. The implemented line search satisfies the Armijo rule which guarantees the local convergence of the method. For more details about this issue the reader can refer to [14]. Besides, a structured strategy is applied to perform the line search. The descent direction is obtained using the gradient of the objective function $f(\mathbf{x})$, in which the free vertex (node) $\mathbf{x}$ is the unknown: $f(\mathbf{x}) = \|K(\mathbf{x})\|_2 = \left[\sum_{m=0}^{M-1} \kappa_m(\mathbf{x})^2\right]^{1/2}$, where κ_m represents the condition number associated to the tetrahedron m, the moving node having M

sub-mesh elements. We then compute the steepest descent $\mathbf{p} = -\nabla f$ and find the position which gives minimum $f(\mathbf{x})$.

3 DD-Coupling

The Chimera method can be viewed as an overlapping domain decomposition technique, where transmission conditions are imposed on the interfaces of the subdomains, see [17]. A key point of the Chimera method is the way the information on the artificial boundaries is transferred, that is, the coupling. The different classical options depends on the type of the transmission conditions imposed on the interfaces. The most typical are Dirichlet/Dirichlet (D/D) coupling, also known as Schwarz' method, Dirichlet/Neumann (D/N) coupling, Dirichlet/Robin (D/R) coupling, Robin/Robin(R/R) coupling. In the litterature, the coupled system is usually solved iteratively. In each subdomain Ω_i local problems are solved by using as boundary conditions (of Dirichlet or Robin type) the values form its neighbours Ω_j until convergence is achieved. Relaxation is often needed to obtain this convergence and depends on the local character of the equation. In [8], the equivalence between the one-domain formulation and overlapping domain decomposition methods of Dirichlet/Neumann(Robin) type is shown at the continuous level. The equivalence is no longer true at the discrete level.

We have developed in this work a new way of coupling the subdomains that we refer to as Extension-Dirichlet (Ext+D). The advantage of the method is that it is implicit and parallel. Therefore, no additional iterative loop is introduced and a-fortiori the convergence of the method has no relation with the overlap. The idea consists in extending the subdomains from their interfaces to their neighboring subdomains, and imposing the Dirichlet condition implicitly, by connecting their extension to the nodes of the neighbors. This method is equivalent, in practice, to imposing Dirichlet boundary condition and eliminating it.

To illustrate the method, let us solve a diffusion equation, $\Delta u = 0$ using the Galerking method in domain $[0,1]$ discretized in 4 linear elements, with the boundary conditions, $u(0) = 1$ and $u(1) = 3$. The analytical solutions is $u = -2x + 1$. Figure 2 (Left) shows the two unconnected subdomains and the corresponding assembled global matrix. Then, Fig. 2 (Center) shows, for the same example, the results of an implicit Dirichelt/Dirichlet coupling. To achieve this, $u_3 - u_5 = 0$ substitutes line 3 and $u_4 - u_2 = 0$ subsitutes line 4. The (Ext+D)^2 method we propose is illustrated in Fig. 2 (Right). Starting with the matrix of Fig. 2 (Left), we perform the following:

- Extend node 3 shape function to node 6 of the second subdomain. This provides additional terms in the equation for node 3.
- Extend node 4 shape function to node 1 of the second subdomain. This provides additional terms in the equation for node 4.

We can observe that in practice the (Ext+D)^2 method creates new elements. In this example the new elements are 3–6 and 4–1. The element matrices and RHS's are

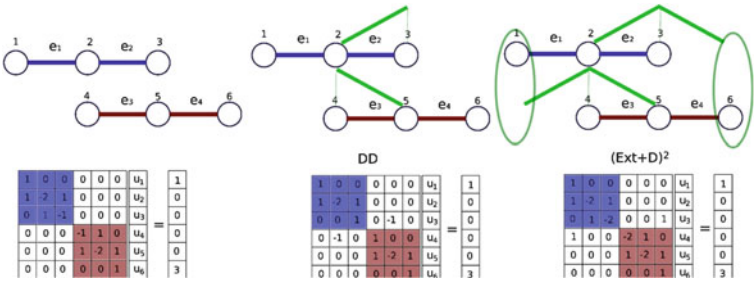


Fig. 2. (*Left*) Problem statement and domain. (*Center*) Dirichlet/Dirichlet assembled. (*Right*) (Ext+D)^2

computed as any other elements of the mesh, but only the lines of node 3 and node 4 of these matrices and RHS's are assembled into the global matrix, respectively.

The main difficulty of the method is to be able to construct a proper extension from one interface node to the other subdomain. This task is specially complex in the 3D case, mainly due to the restriction that the extension must be closed. In variational terms, this means that the extension has a compact support. We are going to describe the way to create the extensions on the interface Γ_{ij} between subdomain Ω_i and subdomain Ω_j in the 2D case. The process, illustrated in Fig. 3, consists in the

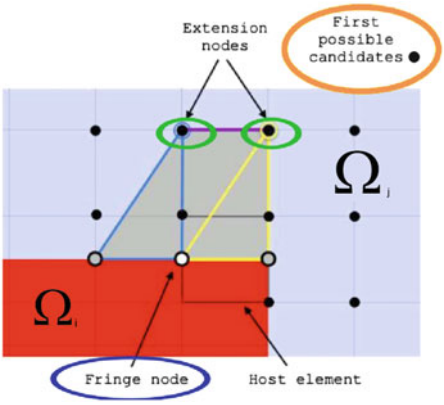


Fig. 3. 2D extensions

following.

- For a fringe node of Ω_i, identify the host element in Ω_j.
- The nodes connected to this host element are the possible candidates to create the triangles that form the associated extension. They are the black nodes.
- Construct two triangles (blue and yellow) connected to the boundaries of the fringe node.

- Close the result with a third one (purple).

The choice of the *extension nodes* (blue and yellow circled) is based on a quality criterion of the resulting triangles [7], among all the possibilities for the previous list. The third node of the triangle is the other node that forms the interface boundary.

4 Numerical Example

Figure 4 shows some results obtained for a flow around a boat. The Navier-Stokes equations are solved together with a level set function and one-equation Spalart-Allmaras turbulence model. The space discretization is a variational multiscale finite element method. The complete description of the algorithm can be found in [10, 11, 15] This complex case computed with 256 CPU's reflects the versatile property of our method and its parallel capacity. The first figure shows the extension elements while the second one the velocity module.

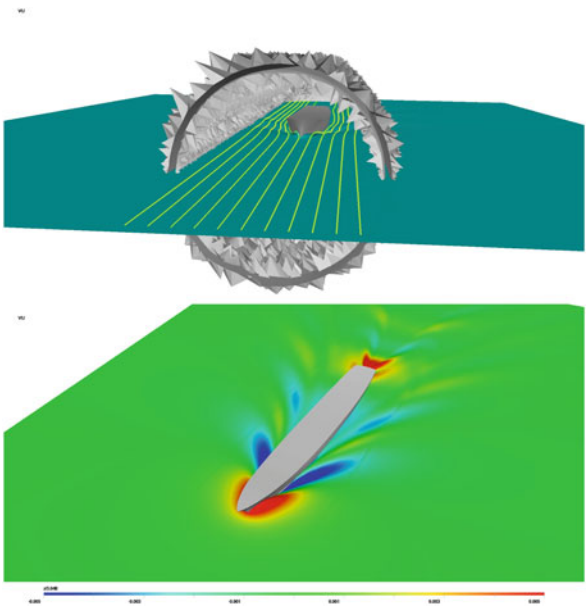


Fig. 4. (*Top*) Extension elements. (*Bottom*) Level set

5 Conclusions

We have devised in this paper a domain decomposition method, referred as (Ext+D)^2 which is based on the explicit construction of extension elements assembled *almost*

as any other element so that the implementation is straightforward. It consists in imposing implicitly Dirichlet transmission conditions and does not introduce any additional iterative loop to the algorithm. Another strength of the method is that it is naturally parallel. However, aspects like conservation should be treated in order to complete the analysis of the method.

Bibliography

[1] J.A. Benek, P.G. Buning, and J.L. Steger. A 3-d chimera grid embedding technique. *AIAA*, 1983.

[2] F. Brezzi, J.L. Lions, and O. Pironneau. Analysis of a chimera method. *C.R.A.S.*, 332:655–660, 2001.

[3] J.J. Chattot and Y. Wang. Improvement treatment of intersecting bodies with the chimera method and validation with a simple fast flow solver. *Computer & Fluids*, 27:721–740, 1998.

[4] J. Dompierre, P. Labbé, F. Guibault, and R. Camarero. Proposal of benchmarks for 3d unstructured tetrahedral mesh optimization. Technical report, Centre de Recherche en Calcul Appliqué, 1998.

[5] L. A. Freitag. On combining laplacian and optimization-based mesh smoothing techniques. In *TRENDS IN UNSTRUCTURED MESH GENERATION*, pages 37–43, 1997.

[6] L.A. Freitag and P.M. Knupp. Tetrahedral element shape optimization via the jacobian determinant and condition number. In *Proceedings of the 8th International Meshing Roundtable, South Lake Tahoe, CA, Sandia National Laboratories, Albuquerque, USA*, pages 247–258. Citeseer, 1999.

[7] P.L. George. Improvements on delaunay-based three-dimensional automatic mesh generator. *Finite Elements in Analysis and Design*, 25:297–317, 1997.

[8] G. Houzeaux. *A Geometrical Domain De Method in Computational Fluid Dynamics*. PhD thesis, CIMNE, 2002.

[9] G. Houzeaux and R. Codina. A chimera method based on a dirichlet/neumann(robin) coupling for the navier-stokes equations. *Comput. Methods Appl. Mech. Engrg.*, 192:3343–3377, 2003.

[10] G. Houzeaux and J. Principe. A variational subgrid scale model for transient incompressible flows. *Int. J. CFD*, 22(3):135–152, 2008.

[11] G. Houzeaux, R. Aubry, and M. Vázquez. Extension of fractional step techniques for incompressible flows: The preconditioned orthomin(1) for the pressure schur complement. *Computers & Fluids*, 44:297–313, 2011. ISSN 0045–7930. doi: DOI:10.1016/j.compfluid.2011.01.017.

[12] A. Khamayseh and A. Kuprat. Deterministic point inclusion methods for computational applications with complex geometry. *Computational Science & Discovery*, 1, 2008.

[13] S. Lo. A new mesh generation scheme for arbitraryplanar domain. *International Journal for Numerical Methods in Engineering*, 21:1403–1426, 1985.

[14] J. Nocedal and S. J. Wright. *Numerical Optimization*. Springer, 2006.

[15] H. Owen, G. Houzeaux, C. Samaniego, A.C. Lesage, and M. Vázquez. Recent ship hydrodynamics developments in the parallel two–fluid flow solver alya. *Submitted to Computers & Fluids*, 2011.

[16] E. Pärt-Enanader. *Overlapping Grids and Applications in Gas Dynamics*. PhD thesis, Uppsala Universitet, Sweden, 1995.

[17] A. Quarteroni and A. Valli. *Domain decomposition methods for partial differential equations*. Oxford University Press, USA, 1999.

[18] C. Samaniego, G. Houzeaux, and M. Vázquez. A parallel fluid-particle solver. *Congreso de Métodos Numéricos em Engenharia*, 2011.

[19] J.L. Steger and F.C. Dougherty J.A. Benek. A chimera grid scheme. *Advances in Grid GEneration*, 5:59–69, 1983.

[20] J.L. Steger and J.A. Benek. On the use of composite grid schemes in computational aerodynamics. *Comp. Meth. Appl. Mech. Eng.*, 64:301–320, 1987.

[21] G. Taubin. A signal processing approach to fair surface design. Technical report, IBM Research, Febreary 1995.

[22] G. Taubin. Curve and surface smoothing without shrinkage. In *Fifth International Conference on Computer Vision*, 1995.

Optimized Schwarz Waveform Relaxation for Porous Media Applications

Caroline Japhet[1] and Pascal Omnes[2]

[1] LAGA, Université Paris XIII, 99 Avenue J-B Clément, 93430 Villetaneuse, France and CSCAMM, University of Maryland College Park, MD 20742 USA, japhet@math.univ-paris13.fr, partially supported by GdR MoMaS.
[2] CEA, DEN, DM2S-SFME-LSET, F91191 Gif sur Yvette Cedex, France., pascal.omnes@cea.fr.

1 Introduction

Far field simulations of underground nuclear waste disposal involve a number of challenges for numerical simulations: widely differing lengths and time-scales, highly variable coefficients and stringent accuracy requirements. In the site under consideration by the French Agency for Nuclear Waste Management (ANDRA), the repository would be located in a highly impermeable geological layer, whereas the layers just above and below have very different physical properties (see [1]). It is then natural to use different time steps in the various layers, so as to match the time step with the physics. To do this, we propose to adapt a global in time domain decomposition method, based on Schwarz waveform relaxation algorithms, to problems in heterogeneous media. This method has been introduced and analyzed for linear advection-reaction-diffusion problems with continuous coefficients [2, 6] and extended to discontinuous coefficients [3, 4], with asymptotically optimized Robin transmission conditions in [3]. The method is extended to higher dimension in [4] with convergence results and error estimates for rectangular or strip subdomains.

This method is extended to problems with discontinuous porosity in [5]. A new aproach is proposed to determine optimized transmission conditions for domains with highly variable lengths. In this paper we analyse this approach in 1d.

Our model problem for the radionuclide transport is the one dimensional advection-diffusion-reaction equation

$$\begin{aligned} \varphi\partial_t u + a\partial_x u - \partial_x(\nu\partial_x u) + bu = f, \quad & \text{on } \mathbb{R}\times(0,T), \\ u(0,x) = u_0(x), \quad & x\in\mathbb{R}. \end{aligned} \tag{1}$$

We focus on a model problem to show the effect of subdomains with widely differing sizes. We consider a decomposition in $\Omega_1 = (-\infty, 0)$, $\Omega_2 = (0, L)$, $\Omega_3 = (L, \infty)$ with $L << 1$. The reaction coefficient b is taken constant and the coefficients a, ν, and φ are assumed constant on each Ω_k, but may be discontinuous at $x = 0$ and $x = L$,

R. Bank et al. (eds.), *Domain Decomposition Methods in Science and Engineering XX*, 585
Lecture Notes in Computational Science and Engineering 91,
DOI 10.1007/978-3-642-35275-1_69, © Springer-Verlag Berlin Heidelberg 2013

$$\varphi = \varphi_k, \quad a = a_k, \quad \nu = \nu_k, \quad x \in \Omega_k.$$

We introduce the notations

$$\mathscr{L}_k v := \varphi_k \partial_t v + a_k \partial_x v - \partial_x(\nu_k \partial_x v) + bv, \quad \text{on } \Omega_k \times (0,T),$$
$$\boldsymbol{\varphi} := (\varphi_1, \varphi_2, \varphi_3), \boldsymbol{a} := (a_1, a_2, a_3), \boldsymbol{\nu} := (\nu_1, \nu_2, \nu_3).$$

Problem (1) is equivalent to solving problems in subdomains Ω_k

$$\begin{aligned} \mathscr{L}_k u_k &= f, \quad \text{on } \Omega_k \times (0,T), \\ u_k(0,x) &= u_0(x), \quad x \in \Omega_k. \end{aligned}$$

with coupling conditions on the interface $\Gamma_{k,\ell}$ between two neighboring subdomains Ω_k and Ω_ℓ given by

$$u_k = u_\ell, \quad \left(\nu_k \partial_x - a_k\right) u_k = \left(\nu_\ell \partial_x - a_\ell\right) u_\ell, \quad \text{on } \Gamma_{k,\ell} \times (0,T). \tag{2}$$

2 Domain Decomposition Algorithm

A simple algorithm based on relaxation of the coupling conditions (2) does not converge in general (see [7]). Following previous works [2–4], we introduce the Schwarz waveform relaxation algorithm

$$\begin{aligned} \mathscr{L}_k u_k^n &= f, \quad \text{on } \Omega_k \times (0,T), \\ u_k^n(0,x) &= u_0(x), \quad x \in \Omega_k, \\ \left(\nu_k \partial_x - a_k\right) u_k^n + \mathscr{S}_{k,\ell} u_k^n &= \left(\nu_\ell \partial_x - a_\ell\right) u_\ell^{n-1} + \mathscr{S}_{k,\ell} u_\ell^{n-1}, \quad \text{on } \Gamma_{k,\ell} \times (0,T), \end{aligned} \tag{3}$$

where $\mathscr{S}_{k,\ell}$ are linear operators in time and space, defined by

$$\mathscr{S}_{k,\ell} \psi = \bar{p}_{k,\ell} \psi + \bar{q}_{k,\ell} \partial_t \psi.$$

The case $\bar{q}_{k,\ell} = 0$ corresponds to Robin transmission conditions, while the case $\bar{q}_{k,\ell} \neq 0$ corresponds to first order transmission conditions. The well-posedness and convergence have been analyzed for constant porosity in [3] and in higher dimension in [4]. The transmission conditions in (3) imply the coupling conditions (2) at convergence, and lead at the same time to an efficient algorithm, for suitable parameters $\bar{p}_{k,\ell}$ and $\bar{q}_{k,\ell}$ obtained from an optimization of the convergence factor.

Similarly, $\mathscr{S}_{k,\ell}$ are approximations of the best operators related to transparent boundary operators. They can be found using Fourier analysis in the two half-spaces case. This analysis has been done for discontinuous coefficients [3], and in higher dimension and continuous coefficients [2]. The min-max problem has been analysed in one dimension in [3] with asymptotical Robin parameters.

In the field of nuclear waste computations, domains of meter scale are embedded in domains of kilometer scale. The previous optimization of the convergence factor does not take into account the high variability of the domains lengths. Following [5], we determine optimized transmission conditions through the minimization of a convergence factor that takes into account this variability.

2.1 Optimal Transmission Conditions

In order to determine the optimal transmission operators $\mathscr{S}_{k,\ell}$, we compute the convergence factor of the algorithm. Since the problem is linear, we consider the algorithm (3) on the error (i.e. with $f = 0$ and $u_0 = 0$). In order to use a Fourier transform in time, we assume that all functions are extended by 0 for $t < 0$.

Let $e_k^n = u_k^n - u$ be the error in Ω_k at iteration k. The operators $\mathscr{S}_{k,\ell}$ are related to their symbols $\sigma_{k,\ell}(\omega)$ by

$$\mathscr{S}_{k,\ell} u(t) = \frac{1}{2\pi} \int \sigma_{k,\ell}(\omega) \hat{u}(\omega) e^{i\omega t} d\omega.$$

The Fourier transforms $\hat{e}_k^n$ in time of e_k^n are solutions of the ordinary differential equation in the x variable

$$-\nu_k \partial_{xx}^2 \hat{e} + a_k \partial_x \hat{e} + (i\varphi_k \omega + b)\hat{e} = 0.$$

The characteristic roots are

$$r^{\pm}(a_k, \nu_k, \varphi_k, b, \omega) = \frac{a_k \pm \sqrt{d_k}}{2\nu_k}, \quad d_k = a_k^2 + 4\nu_k(i\varphi_k \omega + b). \tag{4}$$

Since $\Re r^+ > 0$, $\Re r^- < 0$, and since we look for solutions which do not increase exponentially in $|x|$, we obtain

$$\begin{aligned} \hat{e}_1^n(x,\omega) &= \alpha_1^n(\omega) e^{r^+(a_1,\nu_1,\varphi_1,b,\omega)x}, \quad \hat{e}_3^n(x,\omega) = \alpha_3^n(\omega) e^{r^-(a_3,\nu_3,\varphi_3,b,\omega)x}, \\ \hat{e}_2^n(x,\omega) &= \alpha_2^n(\omega) e^{r^+(a_2,\nu_2,\varphi_2,b,\omega)x} + \beta_2^n(\omega) e^{r^-(a_2,\nu_2,\varphi_2,b,\omega)x}. \end{aligned} \tag{5}$$

We set $\xi^n = (\alpha_1^n, \alpha_2^n, \beta_2^n, \alpha_3^n)^t$, and $r_k^{\pm} = r^{\pm}(a_k, \nu_k, \varphi_k, b, \omega)$. We define the variables $s_k = s_k(\omega, L)$, $1 \le k \le 8$, by

$$s_1 = \frac{\nu_2 r_2^- - \sigma_{1,2}}{\nu_1 r_1^- - \sigma_{1,2}},\ s_2 = \frac{\nu_2 r_2^+ - \sigma_{1,2}}{\nu_1 r_1^- - \sigma_{1,2}},\ s_3 = \frac{\nu_2 r_2^+ - \sigma_{2,3}}{\nu_2 r_2^- - \sigma_{2,3}} \cdot e^{(r_2^- - r_2^+)L},$$

$$s_5 = \frac{\nu_2 r_2^- + \sigma_{2,1}}{\nu_2 r_2^+ + \sigma_{2,1}},\ s_7 = \frac{\nu_2 r_2^- + \sigma_{3,2}}{\nu_3 r_3^+ + \sigma_{3,2}} e^{(r_2^+ - r_3^-)L},\ s_8 = \frac{\nu_2 r_2^+ + \sigma_{3,2}}{\nu_3 r_3^+ + \sigma_{3,2}} e^{(r_2^- - r_3^-)L},$$

$$s_4 = \frac{\nu_1 r_1^- + \sigma_{2,1}}{\nu_2 r_2^+ + \sigma_{2,1}} \cdot \frac{1}{D},\ s_6 = \frac{\nu_3 r_3^+ - \sigma_{2,3}}{\nu_2 r_2^- - \sigma_{2,3}} \cdot \frac{e^{(r_3^- - r_2^+)L}}{D}, \text{ with } D = s_3 s_5 - 1.$$

We insert (5) into the transmission conditions in (3), and obtain for $n \ge 2$,

$$\xi^n = M \xi^{n-1},$$

where the matrix $M = M(\omega, L)$ is defined by

$$M = \begin{pmatrix} 0 & s_1 & s_2 & 0 \\ s_3 s_4 & 0 & 0 & -s_6 \\ -s_4 & 0 & 0 & s_5 s_6 \\ 0 & s_7 & s_8 & 0 \end{pmatrix}.$$

The convergence factor $\rho(\omega, L)$ for each $\omega \in \mathbb{R}$ is the spectral radius of M.

Remark 1. The choice for the symbols $\sigma_{k,\ell}$

$$\sigma_{1,2} = \nu_2 r_2^+, \quad \sigma_{2,1} = -\nu_1 r_1^-, \quad \sigma_{2,3} = \nu_3 r_3^+, \quad \sigma_{3,2} = -\nu_2 r_2^-, \tag{6}$$

leads to $M^2 = 0$ and thus to optimal convergence in three iterations.

Proposition 1. *The convergence factor is given by*

$$\rho(\omega, L) = \sqrt{\max(|\lambda^-|, |\lambda^+|)},$$

where $\lambda^\pm = \lambda^\pm(\omega, L)$ *is defined by*

$$\lambda^\pm = \frac{\alpha + \beta \pm \sqrt{(\alpha-\beta)^2 + 4\gamma\zeta}}{2},$$

with

$$\alpha = s_1 s_3 s_4 - s_2 s_4,\ \beta = -s_6 s_7 + s_5 s_6 s_8,\ \gamma = s_3 s_4 s_7 - s_4 s_8,\ \zeta = -s_1 s_6 + s_2 s_5 s_6.$$

This follows from the computation of the roots of the characteristic polynomial of M, which is biquadratic. The corresponding operators to (6) are non-local in time. In the next subsection, we therefore approximate the optimal operators by local ones.

2.2 Local Transmission Conditions

We approximate the optimal choice $\sigma_{k,\ell}$ in (6) by polynomials in ω :

$$\sigma_{1,2}^{\text{app}} = \frac{p_{1,2} + a_2}{2} + \frac{q_{1,2}}{2} i\omega, \quad \sigma_{2,1}^{\text{app}} = \frac{p_{2,1} - a_1}{2} + \frac{q_{2,1}}{2} i\omega,$$
$$\sigma_{2,3}^{\text{app}} = \frac{p_{2,3} + a_3}{2} + \frac{q_{2,3}}{2} i\omega, \quad \sigma_{3,2}^{\text{app}} = \frac{p_{3,2} - a_2}{2} + \frac{q_{3,2}}{2} i\omega.$$

In order to simplify the min-max problem, we will consider the following cases for the choice of $p_{k,\ell}$ and $q_{k,\ell}$:

1. (Robin) $p_{k,\ell} = p$, $q_{k,\ell} = 0$,
2. (Zeroth order) $p_{1,2} = p_{3,2} = p_1$, $p_{2,1} = p_{2,3} = p_2$, $q_{k,\ell} = 0$,
3. (First order) $p_{k,\ell} = p$, $q_{k,\ell} = q$,
4. (First order scaled) $p_{k,\ell} = p$, $q_{1,2} = \varphi_2 q$, $q_{2,1} = \varphi_1 q$, $q_{2,3} = \varphi_3 q$, $q_{3,2} = \varphi_2 q$.

Then, the parameters are chosen in order to minimize the convergence factor, i.e. we solve, for $\boldsymbol{p} = p$ in case 1, $\boldsymbol{p} = (p_1, p_2)$ in case 2, and $\boldsymbol{p} = (p, q)$ in cases 3 and 4, the min-max problem

$$\delta_m(L) = \min_{\boldsymbol{p}} \left(\max_{\omega_0 \le \omega \le \omega_{\max}} \rho(\omega, \boldsymbol{p}, \boldsymbol{\varphi}, \boldsymbol{a}, \boldsymbol{\nu}, b, L) \right), \tag{7}$$

where ρ is the spectral radius of M, in which we have replaced $\sigma_{k,\ell}$ by $\sigma_{k,\ell}^{\text{app}}$, and m is the order of the approximation. In numerical computations, the frequencies can be restricted to $\omega_{\max} = \frac{\pi}{\Delta t}$, where Δt is the time step, and $\omega_0 = \frac{\pi}{T}$.

Theorem 1. *We suppose that $a_k = a$, $\varphi_k = \varphi$ and $\nu_k = \nu$, $1 \le k \le 3$, thus $d_k = d$ in (4). Let us consider the Robin case ($\boldsymbol{p} = p$) and the first order case ($\boldsymbol{p} = (p,q)$). Then the convergence factor reduces to*

$$\rho(\omega, \boldsymbol{p}, \varphi, a, \nu, b, L) = \sqrt{\left|\frac{\sigma - \sqrt{d}}{\sigma + \sqrt{d}}\right| \max\left(\left|\frac{\sigma - \mu}{\sigma + \mu}\right|, \left|\frac{\sigma - \eta}{\sigma + \eta}\right|\right)}$$

with

$$\mu = \sqrt{d}\left(\frac{1 + e^{-\frac{\sqrt{d}}{2\nu}L}}{1 - e^{-\frac{\sqrt{d}}{2\nu}L}}\right), \ \eta = \frac{\sqrt{d}}{\mu},$$

and with $\sigma = p$ in the Robin case, and $\sigma = p + qi\omega$ in the first order case. Let $L > 0$ given. Let $\delta_0(L)$ (resp. $\delta_1(L)$) be the solution of (7) for the Robin case (resp. the first order case). For $m = 0$ and $m = 1$, we have $|\delta_m(L)| < 1$.

3 Numerical Results

We use the DG-OSWR method in [4] based on a discontinuous Galerkin method in time, with $\mathbf{P}_1$ finite elements in space in each subdomain. We present an example inspired from nuclear waste simulations, with discontinuous coefficients, and different time and space steps in the subdomains $\Omega_2 = (0.4954, 0.5047)$ (repository), $\Omega_1 = (0, 0.4954)$ and $\Omega_3 = (0.5047, 1)$ (host rock). The parameters for the three subdomains are shown in Table 1. The final time is $T = 0.04$.

	φ	ν	a	b	Δx	Δt
$\Omega_1 \cup \Omega_3$	0.06	0.06	1	0	$5\,10^{-3}$	$T(5\,10^{-3})$
Ω_2	0.1	1	1	0	$5\,10^{-4}$	$T(1\,10^{-3})$

Table 1. Physical and numerical parameters

Let $\boldsymbol{p}_3^\star$ (resp. $\boldsymbol{p}_2^\star$) be the parameters derived from a numerical minimization of the three domains convergence factor in (7) (resp. from the two half-spaces convergence factor in [3]). Figure 1 shows $\rho(\omega, \boldsymbol{p}_3^\star, L)$ (solid line) and $\rho(\omega, \boldsymbol{p}_2^\star, L)$ (dashed line) versus ω for $\Delta t = T(5\ 10^{-3})$. We observe that the solution of (7) is characterized by an equioscillation property (at the star marks), as in the two half-spaces case (see [2]). Moreover, for first order transmission conditions, we see that a scaling with the porosity is important only when the parameters are computed from the two half-spaces analysis.

On Fig. 2 we show the error after 20 iterations when running the algorithm on the discretized problem, with $u_0 = f = 0$ and random initial guess on the interfaces, for various values of the Robin parameter p (left) and the zeroth order parameters p_1, p_2 (right) (in that case, the values obtain with the two half-spaces analysis is not in the range values of the figure).

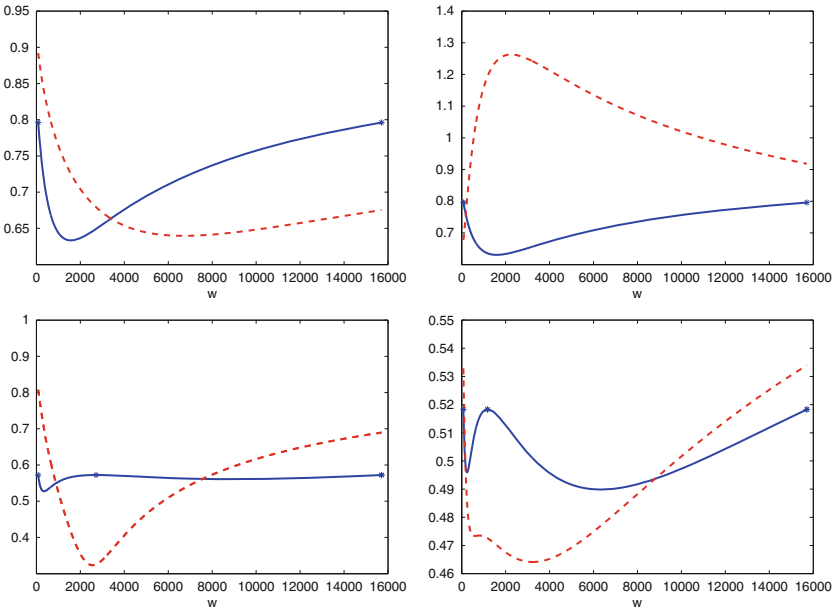


Fig. 1. Convergence factor $\rho(\omega, \boldsymbol{p}_3^*, L)$ (*solid line*) and $\rho(\omega, \boldsymbol{p}_2^*, L)$ (*dashed line*) versus ω: *Top left*: Robin, *top right*: zeroth order, *left bottom*: first order, *right bottom*: first order scaled

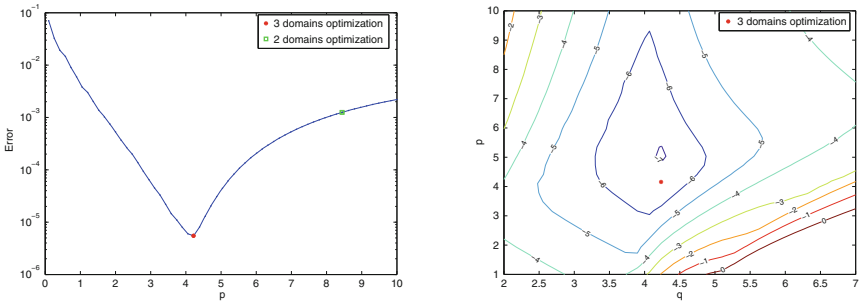


Fig. 2. Error after 20 iterations: *Left*: for various values of the Robin parameter p (the *lower left star marks* $p_3^\star$ whereas the *upper right circle* shows $p_2^\star$), *Right*: the level curves for various values of the zeroth order parameters p_1, p_2 (the *star marks* the parameter $\boldsymbol{p}_3^\star$)

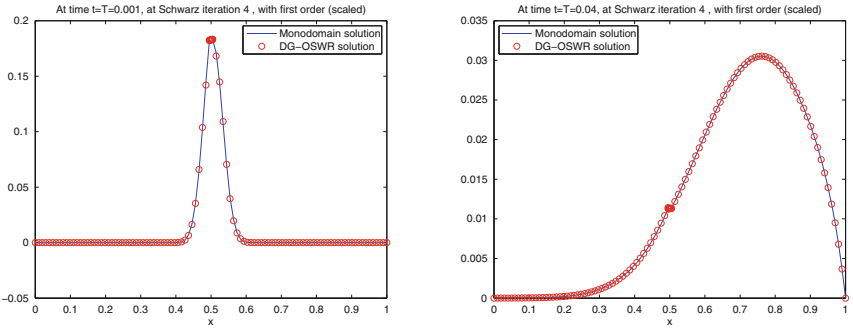


Fig. 3. Evolution of the monodomain solution (*solid line*) and the OSWR solution at iteration 4 (*circle line*): at $t = 0.001$ (*left*), $t = T = 0.04$ (*right*)

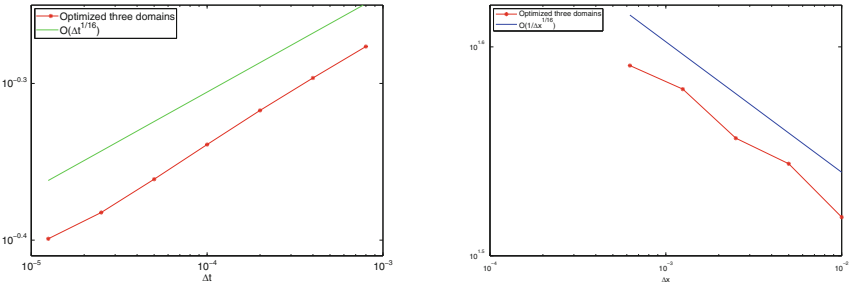


Fig. 4. Asymptotic behavior as the mesh is refined: on the *left* $R(\Delta t)$ and on the *right* where $\Delta t = O(\Delta x)$, the rate for the optimized Schwarz waveform relaxation algorithm with optimized first order (scaled) transmission conditions

On Fig. 3, the solution, with first order (scaled) conditions, at iteration 4 is shown for an initial condition equal to 1 in Ω_2 and 0 elsewhere.

Figure 4 shows, on the left, $R(\Delta t) = 1 - \max_{\pi/T \leq \omega \leq \pi/\Delta t} \rho(\omega, \boldsymbol{p}_3^\star, L)$ versus Δt, i.e. the convergence factor behaves like $1 - O(\Delta t)^{1/16}$, with first order (scaled) optimized transmission conditions. On the right, we run the OSWR algorithm until the error becomes smaller than 10^{-11}, and count the number of iterations. We start with $\Delta t = T/100$ in each subdomain, and repeat this experiment dividing Δx and Δt by 2 several times. We observe that the asymptotic result on the left predicts very well the numerical behavior of the algorithm given on the right.

Bibliography

[1] Stephan Schumacher Alain Bourgeat, Michel Kern and Jean Talandier. The couplex test cases: Nuclear waste disposal simulation. *Computational Geosciences*, 8:83–98, 2004.

[2] D. Bennequin, M.J. Gander, and L.Halpern. A homographic best approximation problem with application to optimized Schwarz waveform relaxation. *Math. Comp.*, 78:185–223, 2009.

[3] M.J. Gander, L. Halpern, and M. Kern. Schwarz waveform relaxation method for advection–diffusion–reaction problems with discontinuous coefficients and non-matching grids. In O.B. Widlund and D.E. Keyes, editors, *Decomposition Methods in Science and Engineering XVI*, volume 55 of *Lecture Notes in Computational Science and Engineering*, pages 916–920. Springer, 2007.

[4] L. Halpern, C. Japhet, and J. Szeftel. Discontinuous Galerkin and nonconforming in time optimized Schwarz waveform relaxation. In O. Widlund Y. Huang, R. Kornhuber and J. Xu, editors, *Domain Decomposition Methods in Science and Engineering XIX*, volume 78 of *Lecture Notes in Computational Science and Engineering*, pages 133–140, 2010.

[5] L. Halpern, C. Japhet, and P. Omnes. Nonconforming in time domain decomposition method for porous media applications. In J. C. F. Pereira and A. Sequeira, editors, *V European Conference on Computational Fluid Dynamics ECCOMAS CFD*, Lisbon, Portugal,14–17 June 2010.

[6] V. Martin. An optimized Schwarz waveform relaxation method for the unsteady convection diffusion equation in two dimensions. *Appl. Numer. Math.*, 52:401–428, 2005.

[7] Alfio Quarteroni and Alberto Valli. *Domain Decomposition Methods for Partial Differential Equations*. Oxford Science Publications, 1997.

On Block Preconditioners for Generalized Saddle Point Problems

Piotr Krzyżanowski

University of Warsaw, ul. Banacha 2, 02-097 Warszawa, Poland;
piotr.krzyzanowski@mimuw.edu.pl

1 Introduction

We consider a symmetric system of linear equations with a block structure,

$$\mathscr{M}\begin{pmatrix} u \\ p \end{pmatrix} \equiv \begin{pmatrix} A & B^T \\ B & -C \end{pmatrix}\begin{pmatrix} u \\ p \end{pmatrix} = \begin{pmatrix} F \\ G \end{pmatrix}. \tag{1}$$

We assume that A is $n \times n$ and C is an $m \times m$ matrix. Many such systems arise from the discretization of (systems of) partial differential equations. For example, Stokes equations discretized with stable finite elements or a mixed finite element method for second order elliptic PDEs lead to a positive definite matrix A and to $C = 0$, so that (1) has a genuine saddle point structure. Certain other PDE problems may result in an indefinite matrix A, or a semidefinite matrix A with a large kernel, which gives (1) the structure of a so called generalized saddle point problem. Linear elasticity equations modelling nearly incompressible materials discretized with mixed finite elements result in both matrices A and C being positive definite, having thus a nature of a penalized saddle point problem. All systems mentioned above have a common feature that the matrix of (1) is indefinite.

The specific structure of (1) makes it possible to design efficient solution methods which intensively exploit the properties of the system, see the recent survey of [4] on the state-of-the-art in this field. Systems derived from the discretization of PDEs are usually very large and sparse, and typically are solved by some iterative method. Unfortunately, these systems are ill-conditioned with respect to the mesh size h, so preconditioning is necessary in order to keep the number of iterations within a reasonable limit. Applying a left preconditioner $\mathscr{P}$, one then solves a problem with a preconditioned matrix $\mathscr{P}^{-1}\mathscr{M}$. We shall consider preconditioners of the form

$$\mathscr{P}_{\mathrm{d}} = \begin{pmatrix} I & \\ cBA_0^{-1} & I \end{pmatrix}\begin{pmatrix} A_0 & \\ & S_0 \end{pmatrix}\begin{pmatrix} I & dA_0^{-1}B^T \\ & I \end{pmatrix} \tag{2}$$

or

$$\mathscr{P}_{\mathrm{p}} = \begin{pmatrix} I & dB^TS_0^{-1} \\ & I \end{pmatrix}\begin{pmatrix} A_0 & \\ & S_0 \end{pmatrix}\begin{pmatrix} I & \\ cS_0^{-1}B & I \end{pmatrix}, \tag{3}$$

R. Bank et al. (eds.), *Domain Decomposition Methods in Science and Engineering XX*, Lecture Notes in Computational Science and Engineering 91,
DOI 10.1007/978-3-642-35275-1_70,

where A_0 and S_0 are symmetric, positive (or negative) definite matrices whose inverses are *easy to apply* and $c,d \in \{-1,+1\}$. In accordance with [8], we will refer to $\mathscr{P}_\mathrm{d}$ as the family of dual block preconditioners and to $\mathscr{P}_\mathrm{p}$ as the family of primal block preconditioners.

Many popular block preconditioners can be formed by choosing appropriate values of c and d in the formulas above. For example, a block diagonal preconditioner, cf. e.g. [2, 6, 9, 13, 19, 21] corresponds to $c = d = 0$ above. Block triangular preconditioners considered e.g. in [7, 14, 22] and the Bramble–Pasciak preconditioner as well, see [5], are obtained with either c or d equal to zero. The choice $c = d = 1$ in (2) produces a symmetric indefinite preconditioner, see [3, 20, 24, 25], while the same choice in (3) leads to a primal based penalty preconditioner, [1, 8].

It is straightforward that solving a system with $\mathscr{P}_\mathrm{d}$ requires one solve with S_0 and at most two solves with A_0, while applying $\mathscr{P}_\mathrm{p}$ to a vector takes one solve with A_0 and at most two solves with S_0. When $cd = 0$, both types of preconditioners require only one solve with A_0 and one with S_0.

Let us stress that when (1) arises from finite element discretization of PDEs, there is a possibility to use other than block preconditioning approaches. On the other hand, for many types of discretizations and problems, specialized methods based on direct construction of a multigrid or domain decomposition preconditioner—although usually outperforming block preconditioners, [15]—may take a considerable effort to develop, implement and analyse. Since the block preconditioning approach as discussed here turns out to be based on preconditioners for symmetric positive definite matrices, this property makes it a viable and robust alternative to custom methods, as in this case one can efficiently reuse existing theory and software to solve more complex problems. This feature has been recognized in the software package PETSc, see [23], where a family of so called field-splitting preconditioners has recently been implemented.

2 Eigenvalue Estimates of the Preconditioned System

Eigenvalue clustering is vital for the convergence of a Krylov method, so it is important to bound the spectrum of $\mathscr{P}^{-1}\mathscr{M}$, where $\mathscr{P}$ stands for either $\mathscr{P}_\mathrm{d}$ or $\mathscr{P}_\mathrm{p}$. Inspired by the block nature of the problem, which imposes a decomposition of the unknowns into two parts $(u,p) \in R^n \times R^m$, let us define a block diagonal, symmetric, positive definite matrix

$$\mathscr{J} = \begin{pmatrix} \tilde{A}_0 & \\ & \tilde{S}_0 \end{pmatrix},$$

where $\tilde{A}_0$ is either A_0, if A_0 is positive definite, or $(-A_0)$, if A_0 is negative definite; $\tilde{S}_0$ is defined in the same way. We assume there exist positive constants m_0 and m_1 such that

$$m_0||x||_{\mathscr{J}} \leq ||\mathscr{M}x||_{\mathscr{J}^{-1}} \leq m_1||x||_{\mathscr{J}} \qquad \forall x \in R^n \times R^m,$$

where

$$||\begin{pmatrix} u \\ p \end{pmatrix}||^2_{\mathscr{J}} = ||u||^2_{\tilde{A}_0} + ||p||^2_{\tilde{S}_0},$$

This is nothing but a stability and continuity assumption in an appropriate norm, see also [18]. At the same time we suppose there exists a constant b_0 such that for any $u \in R^n$ and $p \in R^m$,

$$|p^T Bu| \leq b_0 ||u||_{\tilde{A}_0} ||p||_{\tilde{S}_0}.$$

Finally, we assume that for some $\delta \in \{-1, +1\}$, the matrix $\mathscr{H}$ is positive definite, where $\mathscr{H}$ is equal to either $\mathscr{H}_\mathrm{d}$ or $\mathscr{H}_\mathrm{p}$ (depending on whether we are addressing $\mathscr{P}_\mathrm{d}$ or $\mathscr{P}_\mathrm{p}$), with

$$\mathscr{H}_\mathrm{d} = \delta \begin{pmatrix} A_0 - cA & \\ & S_0 + cdBA_0^{-1}B^T + dC \end{pmatrix},$$

$$\mathscr{H}_\mathrm{p} = \delta \begin{pmatrix} A_0 + cdB^T S_0^{-1} B - cA & \\ & S_0 + dC \end{pmatrix}.$$

It turns out that then both $\mathscr{H}_\mathrm{d}{\mathscr{P}_\mathrm{d}}^{-1}\mathscr{M}$ and $\mathscr{H}_\mathrm{p}{\mathscr{P}_\mathrm{p}}^{-1}\mathscr{M}$ are symmetric and the eigenvalues of the preconditioned matrix are bounded as stated in the following theorem, whose proof appeared in [16]:

Theorem 1. *Suppose the above assumptions are fulfilled. If λ is an eigenvalue of ${\mathscr{P}_\mathrm{d}}^{-1}\mathscr{M}$ or of ${\mathscr{P}_\mathrm{p}}^{-1}\mathscr{M}$, then it is real and satisfies*

$$\frac{m_0}{2(1+b_0^2)} \leq |\lambda| \leq 2m_1(1+b_0^2).$$

Let us mention that earlier Klawonn [12] proved a similar result for block diagonal preconditioning matrices.

2.1 Example Application: Stabilized Stokes Equations

Theorem 1 relies on the stability of (1) and therefore indicates that block preconditioners can be used also in the case when the inf-sup condition is not satisfied and one uses a so called stabilized method. As a model example let us consider a stabilized $Q_1 - Q_1$ discretization of Stokes equations

$$\begin{aligned} -\Delta u + \nabla p &= f, \\ \nabla \cdot u &= 0. \end{aligned}$$

Let $\mathscr{T}_h$ denote a shape-regular, quasi-uniform triangulation of a polygonal $\Omega \subset R^2$ into quadrilaterals. Define the finite dimensional spaces of bilinear finite elements:

$$V_h = \{v \in [H_0^1(\Omega)]^2 : v_{|_\kappa} \in [Q_1(\kappa)]^2 \quad \forall \kappa \in \mathscr{T}_h\}$$

and

$$W_h = \{q \in L_0^2(\Omega) \cap C(\Omega) : q_{|_\kappa} \in Q_1(\kappa) \quad \forall \kappa \in \mathscr{T}_h\},$$

where $Q_1(\kappa)$ denotes the space of bilinear functions on κ. Since V_h and W_h do not satisfy the inf-sup condition the following stabilized discretization has been introduced in [11]:

$$\begin{cases} (\nabla u_h, \nabla v_h)_{L^2(\Omega)} - (\mathrm{div} v_h, p_h)_{L^2(\Omega)} = (f, v_h)_{L^2(\Omega)} & \forall v_h \in V_h, \\ -(\mathrm{div}\, u_h, q_h)_{L^2(\Omega)} - c(p_h, q_h) = -\tau \sum_{\kappa \in \mathscr{T}_h} h_\kappa^2 (f, \nabla q_h)_{L^2(\kappa)} & \forall q_h \in W_h, \end{cases} \tag{4}$$

where

$$c(p_h, q_h) = \tau \sum_{\kappa \in \mathscr{T}_h} h_\kappa^2 (\nabla p_h, \nabla q_h)_{L^2(\kappa)}$$

and $\tau > 0$ is some prescribed parameter, independent of h. As the above system is stable and continuous in the norm $\left(||u||^2_{H^1_0} + ||p||^2_{L^2} \right)^{1/2}$, one concludes that an optimal preconditioner (with respect to the mesh size h) can be obtained with either $\mathscr{P}_\mathrm{d}$ or $\mathscr{P}_\mathrm{p}$, where $\tilde{A}_0$ is spectrally equivalent to the discrete Lapacian operator and $\tilde{S}_0$ is spectrally equivalent to the pressure mass matrix. These operators may require some pre-scaling in order to make either $\mathscr{H}_\mathrm{d}$ or $\mathscr{H}_\mathrm{d}$ positive definite.

Numerical Experiments

We confirm the above findings running experiments for a stabilized $Q_1 - Q_1$ discretization of the Stokes system on a unit square, obtained under MATLAB with the software package IFISS 2.2, see [10].

We investigated the number of iterations of the preconditioned conjugate residual method required to reduce the residual norm by a factor of 10^6. We experimented with $\mathscr{P}_\mathrm{d}$ having one of the following forms: block diagonal ($c = 1, d = 0$), upper triangular ($c = 0, d = 1$) and lower triangular ($c = d = 0$) (see [17] for implementation details) for varying mesh size h. The results for the case when $A_0 = A$ and $S_0 = M$ (as suggested by the above analysis) are provided below, confirming a convergence rate independent of h:

$n+m$	243	867	3,267	12,675	49,923
Lower triangular	17	21	21	22	23
Upper triangular	16	16	16	16	16
Diagonal	32	35	37	39	39

In order to show a more realistic choice of A_0, we used A_0^{-1} defined by means of the incomplete Cholesky factorization of A, with drop tolerance 10^{-3}. Since for our model problem the quality of the incomplete Cholesky factorization degrades slowly with increasing size of the system, this is also reflected in an increase of the iteration counts:

$n+m$	243	867	3,267	12,675	49,923
Lower triangular	18	20	24	35	113
Upper triangular	17	17	20	33	—
Diagonal	33	38	48	74	132

It has been observed that (at least in our implementation) the best solution times were obtained mostly for triangular preconditioners.

3 Conclusions

We have presented two classes of block preconditioners for symmetric saddle point problems and provided eigenvalue estimates of the preconditioned system $\mathscr{P}^{-1}\mathscr{M}$ under a quite general assumption of the stability and continuity of the problem being solved. In the context of PDEs, based upon this result, an iterative method, optimal with respect to the mesh size h, can be designed, which may reuse existing state-of-the-art preconditioners or fast solvers for certain elliptic problems.

Acknowledgments The research has been partially supported in part by Polish Ministry of Science and Higher Education grant N N201 0069 33.

Bibliography

[1] O. Axelsson. Preconditioning of indefinite problems by regularization. *SIAM J. Numer. Anal.*, 16(1):58–69, 1979.

[2] Owe Axelsson and Maya Neytcheva. Eigenvalue estimates for preconditioned saddle point matrices. *Numer. Linear Algebra Appl.*, 13(4):339–360, 2006.

[3] Randolph E. Bank, Bruno D. Welfert, and Harry Yserentant. A class of iterative methods for solving saddle point problems. *Numer. Math.*, 56(7):645–666, 1990.

[4] Michele Benzi, Gene H. Golub, and Jörg Liesen. Numerical solution of saddle point problems. *Acta Numer.*, 14:1–137, 2005.

[5] James H. Bramble and Joseph E. Pasciak. A preconditioning technique for indefinite systems resulting from mixed approximations of elliptic problems. *Math. Comp.*, 50(181):1–17, 1988.

[6] J.H. Bramble and J.E. Pasciak. Iterative techniques for time dependent Stokes problems. *Comput. Math. Appl.*, 33(1–2):13–30, 1997.

[7] Zhi-Hao Cao. Positive stable block triangular preconditioners for symmetric saddle point problems. *Appl. Numer. Math.*, 57(8):899–910, 2007.

[8] C. R. Dohrmann and R. B. Lehoucq. A primal-based penalty preconditioner for elliptic saddle point systems. *SIAM J. Numer. Anal.*, 44(1):270–282 (electronic), 2006.

[9] E. G. D'yakonov. Iterative methods with saddle operators. *Dokl. Akad. Nauk SSSR*, 292(5):1037–1041, 1987.

[10] Howard C. Elman, Alison Ramage, and David J. Silvester. Algorithm 886: IFISS, a Matlab toolbox for modelling incompressible flow. *ACM Trans. Math. Software*, 33(2):Art. 14, 18, 2007. software available at http://www.cs.umd.edu/~elman/ifiss.html.

[11] L.P. Franca, T.J.R. Hughes, and R. Stenberg. Stabilised finite element methods for the Stokes problem. In R.A. Nicolaides and M.D. Gunzburger, editors, *Incompressible Computational Fluid Dynamics – Trends and Advances*, pages 87–107, London, 1993. Cambridge University Press.

[12] Axel Klawonn. *Preconditioners for Indefinite Problems*. PhD thesis, Universität Münster, Germany, 1996.

[13] Axel Klawonn. An optimal preconditioner for a class of saddle point problems with a penalty term. *SIAM J. Sci. Comput.*, 19(2):540–552 (electronic), 1998.

[14] Axel Klawonn. Block-triangular preconditioners for saddle point problems with a penalty term. *SIAM J. Sci. Comput.*, 19(1):172–184 (electronic), 1998. Special issue on iterative methods (Copper Mountain, CO, 1996).

[15] Axel Klawonn and Luca F. Pavarino. A comparison of overlapping Schwarz methods and block preconditioners for saddle point problems. *Numer. Linear Algebra Appl.*, 7(1):1–25, 2000.

[16] Piotr Krzyżanowski. Block preconditioners for saddle point problems resulting from discretizations of partial differential equations. In Owe Axelsson and Janos Karatson, editors, *Efficient preconditioned solution methods for elliptic partial differential equations*. Bentham Science Publishers, 2011. URL www.benthamscience.com/ebooks/9781608052912. E-book available online.

[17] Piotr Krzyżanowski. On block preconditioners for saddle point problems with singular or indefinite (1, 1) block. *Numer. Linear Algebra Appl.*, 18(1):123–140, 2011.

[18] Kent-Andre Mardal and Ragnar Winther. Uniform preconditioners for the time dependent Stokes problem. *Numer. Math.*, 98(2):305–327, 2004.

[19] Torgeir Rusten and Ragnar Winther. A preconditioned iterative method for saddlepoint problems. *SIAM J. Matrix Anal. Appl.*, 13(3):887–904, 1992. Iterative methods in numerical linear algebra (Copper Mountain, CO, 1990).

[20] Joachim Schöberl and Walter Zulehner. Symmetric indefinite preconditioners for saddle point problems with applications to PDE-constrained optimization problems. *SIAM J. Matrix Anal. Appl.*, 29(3):752–773 (electronic), 2007.

[21] David Silvester and Andrew Wathen. Fast iterative solution of stabilised Stokes systems. II. Using general block preconditioners. *SIAM J. Numer. Anal.*, 31(5): 1352–1367, 1994.

[22] V. Simoncini. Block triangular preconditioners for symmetric saddle-point problems. *Appl. Numer. Math.*, 49(1):63–80, 2004.

[23] B.F. Smith, W.D. Gropp, and L.C. McInnes. PETSc 2.0 users manual. Technical Report ANL-95/11, Argonne National Laboratory, 1997. Also available via ftp://www.mcs.anl/pub/petsc/manual.ps.

[24] Panayot S. Vassilevski and Raytcho D. Lazarov. Preconditioning mixed finite element saddle-point elliptic problems. *Numer. Linear Algebra Appl.*, 3(1): 1–20, 1996.

[25] Walter Zulehner. Analysis of iterative methods for saddle point problems: a unified approach. *Math. Comp.*, 71(238):479–505 (electronic), 2002.

Optimal Control of the Convergence Rate of Schwarz Waveform Relaxation Algorithms

Florian Lemarié[1], Laurent Debreu[2], and Eric Blayo[3]

[1] Institute of Geophysics and Planetary Physics, University of California at Los Angeles, 405 Hilgard Avenue, Los Angeles, CA 90024-1567, United States, florian@atmos.ucla.edu

[2] INRIA Grenoble Rhône-Alpes, Montbonnot, 38334 Saint Ismier Cedex, France and Jean Kuntzmann Laboratory, BP 53, 38041 Grenoble Cedex 9, France, laurent.debreu@imag.fr

[3] University of Grenoble and Jean Kuntzmann Laboratory, BP 53, 38041 Grenoble Cedex 9, France, eric.blayo@imag.fr

Summary. In this study we present a *non-overlapping Schwarz waveform relaxation method* applied to the one dimensional unsteady diffusion equation. We derive efficient interface conditions using an *optimal control* approach once the problem is discretized. Those conditions are compared to the usual optimized conditions derived at the PDE level by solving a *min-max problem*. The performance of the proposed methodology is illustrated by numerical experiments.

1 Introduction

Schwarz-like domain decomposition methods are very popular in mathematics, computational sciences, and engineering notably for the implementation of coupling strategies. This type of method, originally introduced for stationary problems, can be extended to evolution problems by adapting the waveform relaxation algorithms to provide the so-called Schwarz waveform relaxation method [2, 4]. The idea behind this method is to separate the spatial domain, over which the time-evolution problem is defined, into subdomains. The resulting time-dependent problems are then solved separately on each subdomains. An iterative process with an exchange of boundary conditions at the interface between the subdomains is then applied to achieve the convergence to the solution of the original problem. To accelerate the convergence speed of the iterative process, it is possible to derive efficient interface conditions by solving an optimization problem related to the convergence rate of the method [e.g.; 1, 5].

In this study, we specifically address the optimization problem arising from the use of *Robin* type transmission conditions in the framework of a *non-overlapping Schwarz waveform relaxation*. For this type of problem, the existing work has been achieved mainly at the *PDE level*, giving rise to the optimized Schwarz waveform

R. Bank et al. (eds.), *Domain Decomposition Methods in Science and Engineering XX*, Lecture Notes in Computational Science and Engineering 91,
DOI 10.1007/978-3-642-35275-1_71,

relaxation algorithm [1, 2, 5]. The objective here is to use the *optimal control theory* paradigm [9] to find parameters optimized at the *discrete level*, and thus to systematically make a comparison with the parameters determined at the PDE level. This paper is organized as follows : in Sect. 2 we briefly recall the basics of optimized Schwarz methods in the framework of a time evolution problem. Section 3 is dedicated to the determination of the *optimal control problem* that we intend to address. Finally, in Sect. 4 we apply our approach to a diffusion problem.

2 Optimization of the Convergence at the PDE Level

2.1 Model Problem and Optimized Schwarz Methods

Let us consider Ω a bounded open set of $\mathbb{R}$. The model problem is to find u such that u satisfies over a time period $[0,T]$

$$\mathscr{L}u = f, \qquad \text{in } \Omega \times [0,T], \tag{1}$$
$$\mathscr{B}u = g, \qquad \text{on } \partial\Omega \times [0,T], \tag{2}$$

where $\mathscr{L}$ and $\mathscr{B}$ are two partial differential operators, and f the forcing. This problem is complemented by an initial condition

$$u(x,0) = u_0(x), \qquad x \in \Omega. \tag{3}$$

We consider a splitting of the domain Ω into two *non-overlapping domains* Ω_1 and Ω_2 communicating through their common interface Γ. The operator $\mathscr{L}$ introduced previously is split into two operators $\mathscr{L}_j$ restricted to Ω_j ($j = 1,2$). By noting $\mathscr{F}_1$, $\mathscr{F}_2$, $\mathscr{G}_1$ and $\mathscr{G}_2$ the operators defining the interface conditions, the alternating form of the *Schwarz waveform relaxation algorithm* reads

$$\left\{\begin{array}{rll} \mathscr{L}_1 u_1^k = f_1, & \text{in } \Omega_1 \times [0,T], \\ u_1^k(x,0) = u_o(x), & x \in \Omega_1, \\ \mathscr{B}_1 u_1^k(x,t) = g_1, & \text{in } [0,T] \times \partial\Omega_1, \\ \mathscr{F}_1 u_1^k(0,t) = \mathscr{F}_2 u_2^{k-1}(0,t), & \text{in } \Gamma \times [0,T], \end{array}\right. \quad \left\{\begin{array}{rll} \mathscr{L}_2 u_2^k = f_2, & \text{in } \Omega_2 \times [0,T], \\ u_2^k(x,0) = u_o(x), & x \in \Omega_2, \\ \mathscr{B}_2 u_2^k(x,t) = g_2, & \text{in } [0,T] \times \partial\Omega_2, \\ \mathscr{G}_2 u_2^k(0,t) = \mathscr{G}_1 u_1^k(0,t), & \text{in } \Gamma \times [0,T], \end{array}\right. \tag{4}$$

where $k = 1,2,\ldots$ is the iteration number, and the initial guess $u_2^0(0,t)$ must be given. The operators $\mathscr{F}_j$ and $\mathscr{G}_j$ must be chosen to impose the desired consistency of the solution on the interface Γ. We consider here the one-dimensional diffusion equation with constant (possibly discontinuous) diffusion coefficients κ_j ($\kappa_j > 0, j = 1,2$). We define $\mathscr{L}_j = \partial_t - \kappa_j \partial_x^2$, $\Omega_1 = (-L_1, 0)$, $\Omega_2 = (0, L_2)$ ($L_1, L_2 \in \mathbb{R}^+$), and $\Gamma = \{x = 0\}$. In this context, we require the equality of the subproblems solutions and of their normal fluxes on the interface Γ,

$$u_1(0,t) = u_2(0,t), \qquad \kappa_1 \partial_x u_1(0,t) = \kappa_2 \partial_x u_2(0,t), \qquad t \in [0,T]. \tag{5}$$

To obtain such a consistency we use mixed boundary conditions of *Robin* type

$$\mathscr{F}_j = -\kappa_j \partial_x + p_1, \qquad \mathscr{G}_j = \kappa_j \partial_x + p_2, \qquad (j = 1,2),$$

where p_1 and p_2 are two parameters that can be optimally chosen to improve the convergence speed of the Schwarz method. Algorithm (4) with two-sided Robin conditions (i.e. for $p_1 \neq p_2$) is well-posed for any choice of p_1 and p_2 such that $p_1 + p_2 > 0$. This result can be shown using a priori energy estimates, as described in [4].

2.2 Optimization of the Convergence Factor

To demonstrate the convergence of algorithm (4) a classical approach [e.g. 6] is to define the error e_j^k between the exact solution $u^\star$ and the iterates u_j^k. A Fourier analysis enables the transformation of the original PDEs into ODEs that can be solved analytically. The analytical solution on each subdomain is then used to define a convergence factor ρ of the corresponding *Schwarz algorithm*. For a diffusion problem, defined on subdomains of infinite size (i.e. assuming $L_1, L_2 \to \infty$), we get

$$\rho(p_1, p_2, \omega) = \left| \frac{(p_2 - \sqrt{i\omega\kappa_2})}{(p_2 + \sqrt{i\omega\kappa_1})} \frac{(p_1 - \sqrt{i\omega\kappa_1})}{(p_1 + \sqrt{i\omega\kappa_2})} \right|, \tag{6}$$

where p_1 and p_2 are two degrees of freedom which can be tuned to accelerate the convergence speed. In (6), $i = \sqrt{-1}$, and $\omega \in \mathbb{R}$ is the angular frequency arising from a Fourier transform in time on e_j^k. A general approach to choose the Robin parameters p_1 and p_2 is to solve a minimax problem [2]

$$\min_{p_1, p_2 \in \mathscr{R}} \left(\max_{\omega \in [\omega_{\min}, \omega_{\max}]} \rho(p_1, p_2, \omega) \right). \tag{7}$$

Because we work in practice on a discrete problem the frequencies allowed by the temporal grid range from $\omega_{\min} = \pi/T$ to $\omega_{\max} = \pi/\Delta t$, where Δt is the time step of the temporal discretization. For the diffusion problem under consideration here, the analytical solution of the optimization problem (7) has been derived in [8] in a general *two-sided* case (i.e. with $p_1 \neq p_2$) with discontinuous coefficients $\kappa_1 \neq \kappa_2$. For the sake of simplicity, we consider in the present study the continuous case ($\kappa_1 = \kappa_2 = \kappa$) and we recall the result found in [8] in this case.

Theorem 1. *Under the assumption $\kappa_1 = \kappa_2 = \kappa$, the optimal parameters $p_1^\star$ and $p_2^\star$ of the minmax problem (7) are given by*

$$p_1^\star = \frac{\alpha\sqrt{2\kappa}}{4}\left[\sqrt{8+v^2} - v\right], \qquad p_2^\star = \frac{\alpha\sqrt{2\kappa}}{4}\left[\sqrt{8+v^2} + v\right],$$

where $\alpha = (\omega_{\min}\omega_{\max})^{1/4}$, $\beta = \alpha^{-1}(\sqrt{\omega_{\min}} + \sqrt{\omega_{\max}})$ and

$$v = \begin{cases} 2\sqrt{\beta - 1} & \text{if } \beta \geq 1 + \sqrt{5}, \\ \sqrt{2\beta^2 - 12} & \text{if } \sqrt{6} \leq \beta < 1 + \sqrt{5}, \\ 0 & \text{if } 2 < \beta < \sqrt{6}. \end{cases}$$

It is worth mentioning that even if the diffusion coefficients are continuous the *two-sided* case provides a faster convergence than the *one-sided* case studied in [4] (Fig. 1).

General Remarks :

- The usual methodology to optimize the convergence at the continuous level comes with a few assumptions that may lead to inaccuracies once the problem is discretized. For example, as discussed in [7] (Sect. 5), the *infinite domain assumption* used to determine the convergence factor (6) may lead to appreciable differences in the optimized parameters compared to an approach taking the finiteness of the subdomains into account. We numerically found that the *infinite domain assumption* is valid as long as the *dimensionless Fourier number* $\mathrm{Fo} = \kappa_j/(L_j^2\omega)$ (with L_j the size of subdomain Ω_j) of the problem does not exceed a critical value $\mathrm{Fo}_c = 0.02$.
- The optimization problem (7) aims at minimizing the maximum value of $\rho(p_1, p_2, \omega)$ over the entire interval $[\omega_{\min}, \omega_{\max}]$. This provides a very robust method general enough to deal with the worst case scenario when all the temporal frequencies are present in the error. An even more efficient way to proceed would be to adjust the values of p_1 and p_2 at each iteration so that those parameters are efficiently chosen to "fight" the remaining frequencies in the error.

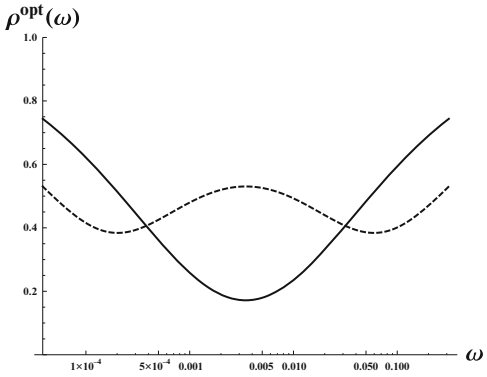


Fig. 1. Convergence factor optimized at the PDE level in the *one-sided* case (*black line*) [4] and in the *two-sided* case (*dashed black line*) [8], for $\kappa = 10^{-2}$ m s^{-1}, $\Delta t = 10$ s, and $T = 2^{13}\Delta t$

3 Optimal Control of the *Robin* Parameters

To investigate the robustness of the optimized parameters once the problem is discretized, the use of the *optimal control theory* appears as a natural choice. We aim at controlling the *Robin* parameter in order to get the best possible convergence speed in the sense of a given cost function $\mathscr{J}$. Moreover, following the approach of [3] and the previous discussion, we consider the possibility to use different parameters p_j for different steps of the iterative process. It is easy to check that by choosing

different parameters at each iteration we still converge to the solution of the global problem. A first way to choose the parameters is to look, at each iteration k, for p_1^k and p_2^k minimizing the error at the interface. In this case the cost function that we intend to minimize at each iteration would be

$$\begin{aligned}\mathscr{J}(p_1^k, p_2^k) = &\frac{w}{2}\int_0^T \left(u_1^k(0,t) - u_2^k(0,t)\right)^2 dt \\ &+ \frac{\widetilde{w}}{2}\int_0^T \left(\kappa_1 \partial_x u_1^k(0,t) - \kappa_2 \partial_x u_2^k(0,t)\right)^2 dt.\end{aligned} \tag{8}$$

The constants w and $\widetilde{w}$ must be chosen to balance both terms, depending on the characteristics of the problem (see Sect. 4). The cost function (8) is designed in agreement with the consistency (5) we want to impose at the interface between subdomains. $\mathscr{J}$ provides a measure of the "inconsistency" of the solution at each iteration k, and is, thus, directly related to the order of magnitude of the errors e_j^k of the algorithm (as shown in Fig. 2). An other strategy could be to minimize the error at a given iteration K. The cost function would thus be

$$\begin{aligned}\mathscr{J}\left((p_1^k, p_2^k)_{k=1,K}\right) = &\frac{w}{2}\int_0^T \left(u_1^K(0,t) - u_2^K(0,t)\right)^2 dt \\ &+ \frac{\widetilde{w}}{2}\int_0^T \left(\kappa_1 \partial_x u_1^K(0,t) - \kappa_2 \partial_x u_2^K(0,t)\right)^2 dt,\end{aligned} \tag{9}$$

leading to an optimization on $2K$ parameters. This latter approach is particularly interesting when we intend to obtain the best possible approximation of the exact solution after a number of iterations set in advance. We propose here to lead our study with this kind of approach with $K = 5$. The *optimal control* approach does not *per se* reduce the computational cost of the algorithm because many evaluations of the cost function are required during the minimization process (see Algorithm 3). We use this approach as a tool to improve our understanding of the behavior of the Robin parameters in order to find new directions to further accelerate the convergence speed when Robin-type interface conditions are used. We denote by $\mathbf{p_1^{\star,num}}$ and $\mathbf{p_2^{\star,num}}$ the parameters found numerically by solving the optimal control problem. Those parameters correspond to two vectors of size K. Similarly we will denote by $\mathrm{p}_1^{\star,\mathrm{ana}}$ and $\mathrm{p}_2^{\star,\mathrm{ana}}$ the parameters found analytically (cf. Theorem 1).

We used Matlab for the computation (Algorithm 3). Note that the well-posedness of the coupling problem (4) is not sufficient to ensure a well-posed optimal control problem. Some additional requirements on the convexity and regularity of the cost function are necessary. We do not provide here such a proof, however we empirically checked that the same solution of the optimal problem is obtained for a wide range of parameter values for the initial guess.

4 Numerical Experiments

We discretized problem (4) using a *backward Euler* scheme in time and a second order scheme defined on a staggered grid in space (see [8] for more details). We

Algorithm 3 Optimal control

```
%== Robin parameters found analytically : p1ana, p2ana
%== Solution of the optimal control problem : p1opt, p2opt
%== Initial guess ==%
x0(1:2:2*K-1)=p1ana;
x0(2:2:2*K  )=p2ana;
%== Solve the optimal control problem ==%
%== the CalcJ function proceeds to K iterations of the
%== Schwarz algorithm using 2K Robin parameters,
%== and computes the associated cost function (9)
x = fminsearch( @CalcJ, x0 );
%== Retrieve the optimized parameters
p1opt(1:K)=x(1:2:2*K-1);
p2opt(1:K)=x(2:2:2*K  );
```

decompose the domain Ω into two non-overlapping subdomains $\Omega_1 = [-H, 0]$ and $\Omega_2 = [0, H]$ with $H = 500$ m. The diffusion coefficient is $\kappa = 10^{-2}$ m^2 s^{-1} and the total simulation time is $T = 2^{13}\Delta t$ with $\Delta t = 10$ s. The parameter values lead to a *dimensionless Fourier number* smaller than 0.02 so that the *infinite domain assumption* is valid. We simulate directly the error equations, i.e. $f_1 = f_2 = 0$ in (4) and $u_0(x) = 0$. We start the iteration with a random initial guess $u_2^0(0,t)$ ($t \in [0,T]$) so that it contains a wide range of the temporal frequencies that can be resolved by the computational grid. This is done to allow a fair comparison as the parameters optimized at the PDE level are optimized assuming that the full range $[\omega_{\min}, \omega_{\max}]$ is present in the error. We first perform the Optimized non-overlapping Schwarz Method (referred as to OSM case) using $\mathrm{p}_1^{\star,\mathrm{ana}}$ and $\mathrm{p}_2^{\star,\mathrm{ana}}$ and then using an optimal control of the *Robin* parameters with $K = 5$ (referred as to OptCon case). We first check that the minimization of cost function $\mathscr{J}$ consistently implies the reduction of the errors $\|e_j\|_\infty$ of the associated algorithm (Fig. 2). For our experiments, we chose $w = 1$ and $\widetilde{w} = H/\kappa$ in (9). We notice that in the OptCon case the convergence speed is significantly improved compared to the OSM case. Indeed, nine iterations of the OSM are required to obtain the same accuracy than the OptCon case after only five iterations. In order to have more insight on the way the parameters $\mathbf{p_1^{\star,num}}$ and $\mathbf{p_2^{\star,num}}$ evolve throughout the iterations we plot, in Fig. 3, the corresponding convergence factor (6) at each iteration. It is striking to realize that the optimal convergence is obtained through a combination of 2-point (equivalent to the *one-sided* case) and 3-point (equivalent to the *two-sided* case) equioscillations sometimes shifted along the ω-axis to adapt to the temporal frequencies still present in the error. The first two iterations aim at working mainly on the high-frequency components while the last three iterations are optimized to work on the low-frequency component. The adaptivity of the *Robin* parameters from one iteration to the other brings more flexibility to the method enabling more scale selectivity.

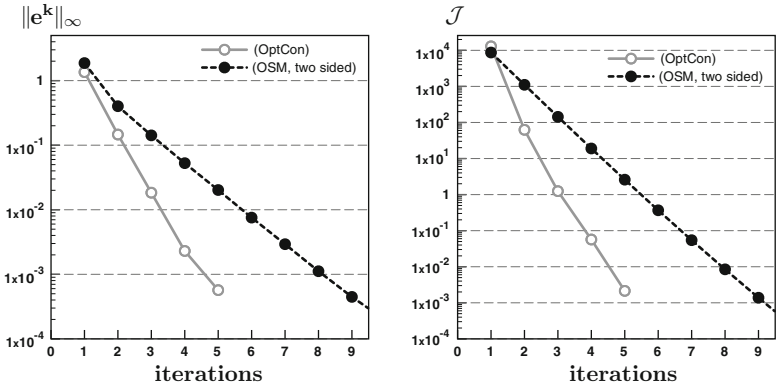


Fig. 2. Evolution of the $\mathscr{L}^\infty$-norm of the error (*left*) and of the cost function $\mathscr{J}$ (*right*) with respect to the iterates k in the OSM and OptCon cases

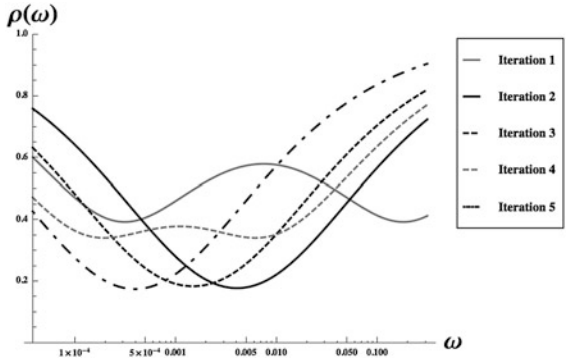


Fig. 3. Sequence of convergence factors $\rho(\omega)$ resulting from the optimal control of the Robin parameters determined to get the best possible convergence after $K = 5$ iterations

5 Conclusion

Due to its simplicity, the use of *Robin-type transmission conditions* is very attractive when one wants to couple unsteady problems defined on non-overlapping subdomains. Once the *Robin* parameters are properly chosen one can achieve a fast convergence [2]. In the present study we showed that there is still room for improvement in the design of the Robin conditions. If the *Robin* parameters are adjusted from one iteration to the other we showed, thanks to an optimal control approach, that we can significantly improve the convergence speed. It is important to emphasize that the *optimal control* paradigm proposed in this study is general enough to be used with any type of PDE and an arbitrary number of subdomains.

Acknowledgments This research was partially supported by the ANR project COMMA (COupling in Multi-physics and multi-scale problems: Models and Algorithms) and by the INRIA project-team MOISE (Modelling, Observation and Identification for Environmental Sciences). We are thankful to Héloïse Pelen (ENS Lyon) for her contribution during her masters internship.

Bibliography

[1] D. Bennequin, M. J. Gander, and L. Halpern. Optimized Schwarz waveform relaxation methods for convection reaction diffusion problems. *Technical Report 2004-24, LAGA, Université Paris 13*, 2004.

[2] M. J. Gander, L. Halpern, and F. Nataf. Optimal convergence for overlapping and non-overlapping Schwarz waveform relaxation. In *Eleventh International Conference on Domain Decomposition Methods (London, 1998)*, pages 27–36 (electronic). DDM.org, Augsburg, 1999.

[3] M. J. Gander and G. H. Golub. A non-overlapping optimized Schwarz method which converges with arbitrarily weak dependence on h. In *Domain decomposition methods in science and engineering*, pages 281–288 (electronic). Natl. Auton. Univ. Mex., México, 2003.

[4] M. J. Gander and L. Halpern. Methodes de relaxation d'ondes pour l'equation de la chaleur en dimension 1. *C. R. Acad. Sci. Paris*, 336(Série I):519–524, 2003.

[5] M. J. Gander and L. Halpern. Optimized Schwarz waveform relaxation methods for advection reaction diffusion problems. *SIAM J. Numer. Anal.*, 45(2):666–697 (electronic), 2007. ISSN 0036–1429. doi: 10.1137/050642137.

[6] M. J. Gander, L. Halpern, and M. Kern. A Schwarz waveform relaxation method for advection-diffusion-reaction problems with discontinuous coefficients and non-matching grids. in *Domain decomposition methods in science and engineering XVI*, vol. 55 of Lect. Notes Comput. Sci. Eng., Springer, Berlin, pp. 283–290, 2007.

[7] F. Lemarié, L. Debreu, and E. Blayo. Optimized global-in-time Schwarz algorithm for diffusion equations with discontinuous and spatially variable coefficients. Research Report RR-6663, INRIA, 2008.

[8] F. Lemarié, L. Debreu, and E. Blayo. Toward an optimized global-in-time Schwarz algorithm for diffusion equations with discontinuous and spatially variable coefficients, part 1 : the constant coefficients case. *Electron. Trans. Numer. Anal.*, 2012. (in revision).

[9] J.-L. Lions. *Contrôle optimal des systèmes gouvernés par des équations aux dérivées partielles*. Dunod, Paris, 1968.

A New Distributed Optimization Approach for Solving CFD Design Problems Using Nash Game Coalition and Evolutionary Algorithms

Jyri Leskinen[1] and Jacques Périaux[12]

[1] University of Jyväskylä, Finland jyri.a.leskinen@gmail.com
[2] CIMNE, Barcelona, Spain jperiaux@gmail.com

1 Introduction

For decades, domain decomposition methods (DDM) have provided a way of solving large-scale problems by distributing the calculation over a number of processing units. In the case of shape optimization, this has been done for each new design introduced by the optimization algorithm. This sequential process introduces a bottleneck.

Shape optimization is often done using gradient-based approaches because of their superior efficiency. Adjoint methods provide a mathematical approach of computing the gradients [4] using calculus of variations. Methods that combine the governing PDEs, their adjoints and shape parameters into one large system of equations are called *one-shot methods* [1, 6]. The optimal shape can be acquired by solving the system of equations only once. Evidently, this approach has several drawbacks. If the objective function is not unimodal, the method does not guarantee capturing the global optimal solution. Also, if the geometry changes are large, mesh deformation is no longer possible and the mesh has to be regenerated which makes this approach costly.

In this paper, a "distributed one-shot" method is introduced. It is based on ideas originating from the fields of game theory, domain decomposition, and evolutionary computing. The aim is to speed up convergence on one hand by decreasing computational time by intelligent parallelism using Nash game strategies and on the other hand by eliminating the bottleneck caused by sequential "state–costate – gradient" chain processing. The evolutionary approach allows the method to be used in global or non-smooth optimization.

1.1 Nash Games in Geometry and Domain Decomposition

Competitive Nash games were introduced by J. Nash [5]. In a competitive game the players maximize their payoff by taking into account the opponents' strategies. Nash games converge into a *Nash equilibrium*. For simplicity, let us consider a two-player

R. Bank et al. (eds.), *Domain Decomposition Methods in Science and Engineering XX*, Lecture Notes in Computational Science and Engineering 91,
DOI 10.1007/978-3-642-35275-1_72,

game. Let S_1 and S_2 be the sets of available strategies of Players 1 and 2 and J_1 and J_2 their payoff functions. A strategy pair $(\bar{x}_1, \bar{x}_2) \in (S_1, S_2)$ is a Nash equilibrium if and only if

$$\begin{aligned} J_1(\bar{x}_1, \bar{x}_2) &= \inf_{x_1 \in S_1} J_1(x_1, \bar{x}_2) \\ J_2(\bar{x}_1, \bar{x}_2) &= \inf_{x_2 \in S_2} J_2(\bar{x}_1, x_2) \end{aligned} \tag{1}$$

The above definition can be easily generalized to a Nash game with N players.

Nash games can also be applied to single-objective optimization. If the objective function J is additively separable, i.e. $J(\mathbf{x}) = \sum_{i=1}^{N} J_i(\mathbf{x}_i)$ and $\min_{\mathbf{x}} J(\mathbf{x}) = \min_{\mathbf{x_i}} \sum_{i=1}^{N} J_i(\mathbf{x}_i) = \mathbf{0}$, a *"virtual" Nash game* can be formed [3]. Since there are no true conflicts between the criteria, the global Nash equilibrium is located at the global optimum.

The Nash approach is well suited for inverse problems. The geometry can often be decomposed into smaller subgeometries which can be optimized concurrently [11]. Similarly, a domain decomposition problem for solving a partial differential equation can be considered as an inverse problem with a Nash game approach where the objective function is to minimize the discrepancy between the local overlapped subdomain solutions,

$$\begin{aligned} JF_1(g_1, \bar{g}_2) &= \int_{\Omega_{1,2}} |\varphi_1(g_1, \bar{g}_2) - \varphi_2(g_1, \bar{g}_2)|^2 \\ JF_2(\bar{g}_1, g_2) &= \int_{\Omega_{1,2}} |\varphi_2(\bar{g}_1, g_2) - \varphi_1(\bar{g}_1, g_2)|^2 \end{aligned} \tag{2}$$

where $|\cdot|$ is the L^2 norm, φ_i is the solution in the subdomain Ω_i and g_i is the vector of values of φ_i on the subdomain interface boundary $\Gamma_{i,j}$. $\Omega_{1,2}$ is the overlapping region (cf. Fig. 1).

In [3, 7], a hierarchical leader–follower Stackelberg game consisting of a pair of Nash games was implemented for nozzle shape reconstruction. The shape players reconstructed the target geometry using a "leader" Nash game, and the flow players reconstructed the flow using a "follower" Nash game. For each new geometry candidate produced by the shape players, a Nash game was run between the flow players. In this paper, a new Nash evolutionary approach is introduced. It replaces the computationally expensive hierarchical game by a single parallel global Nash game coalition.

1.2 Global Nash Game Coalition Algorithm (GNGCA)

The proposed method operates as follows. The geometry of the configuration is divided into subgeometries allocated to shape players whose task is to optimize the shape (or reconstruct the target geometry). Similarly, the flow players minimize the deviation of local solutions on the overlapped region of subdomains. Each shape and flow player evaluate deviation of local solutions or shape optimization with his own Evolutionary Algorithm (EA). After some frequency period, for example a single generation, shape and flow players exchange the elite values among each other. This means the flow is reconstructed along with the geometry making this a "distributed one-shot" method.

This new method is inherently parallel and therefore especially suitable for distributed parallel environments. At the higher level, the flow and shape players operate separately. Depending on the methods used, the optimization process can also be distributed. If an optimizer is used in flow reconstruction, it too can be parallelized. By reducing dimensionality of the geometry problem, algorithmic convergence can be significantly improved. For example, in the case of multi-modal problems splitting the territory can reduce the number of local optima. However, the efficiency of virtual Nash approach is highly dependent on the selected geometry decomposition. Non-optimal splitting can lead in reduced efficiency of the algorithm [11].

2 Test Case Description

The method is validated using a simple position reconstruction problem from the field of computational fluid dynamics. The geometry of the problem consists of a large disk element (radius $\frac{1}{2}$ units) surrounded by $N \geq 2$ smaller disk elements (radii $\frac{1}{8}$ units). The smaller elements are allowed to move in an area constrained by the number of elements: using radial coordinates, $r_k = 2.0^{+0.5}_{-1.3675}$ and $\theta_k = -k\frac{2\pi}{N} - \frac{\pi}{N} \pm \frac{\pi}{4N}$ (see Fig. 1).

This geometry allows the study of a wide variety of different domain and geometry decompositions (cf. Fig. 1 for a 3 element case). The test case can be made more challenging for example by deforming the shapes of the elements. In this paper, 2 and 6 element cases were studied.

The flow is described by the steady compressible potential flow,

$$\begin{aligned} \nabla \cdot \rho \nabla \varphi_k &= 0 \quad \text{in } \Omega_k \\ \varphi_k &= \mathbf{v}_\infty \text{ on } \Gamma_\infty \\ \frac{\partial \varphi_k}{\partial \mathrm{n}} &= 0 \quad \text{on } \Gamma_{1,\ldots,n} \\ \varphi_k &= \varphi_j \text{ on } \Gamma_j \\ \varphi_k &= \varphi_\ell \text{ on } \Gamma_\ell \end{aligned} \tag{3}$$

where k is the index of the subdomain, and j, ℓ the right and left side neighbor domain indexes. Free-flow velocity $\mathbf{v}_\infty = (v_x, v_y) = (v_\infty \cos\alpha, v_\infty \sin\alpha)$, $|v_\infty| = 1$. The angle of attack $\alpha = 0.0^\circ$. The density ρ is calculated using the formula $\rho = \left\{1 + \frac{\gamma-1}{2} M_\infty^2 \left(1 - |v|^2\right)\right\}^{\frac{1}{\gamma-1}}$. The constant $\gamma = 1.4$ is the ratio of specific heats for air. With a free flow Mach number $M_\infty = .3$ the flow is subsonic in the whole domain.

The objective is to reconstruct the original positions of the elements by minimizing the L^2 norm of pressure difference between the computed and target surface pressures: $JS_k(\mathbf{x}_k) = \frac{1}{n_{p_k}} \sum_{i=1}^{n_{p_k}} \left| p_{k_i} - p_{k_i}^{target} \right|^2$ where $\mathbf{x}_k = (r_k, \theta_k)$ is the decomposed design vector and n_{p_k} is the number of pressure points in the region of the decomposed geometry. The vector p_k includes the relevant surface pressure values. The global objective function is the sum of local functions.The objective function for the flow players is the L^2 norm of the discrepancy on the overlapped subregion (Eq. 2).

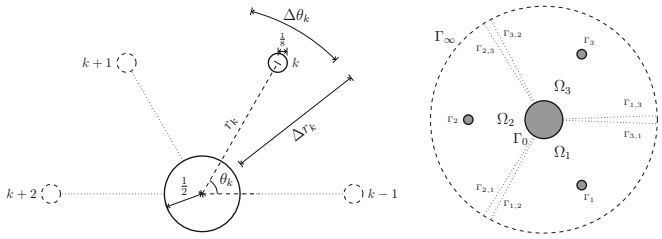


Fig. 1. Test case geometry and example decomposition

3 Test Setting

A variant of the popular Differential Evolution (DE) algorithm is used as the optimization platform. The algorithm, differential evolution with adaptive control parameters (jDE) is described in detail in the original paper [2]. The difference compared to the standard differential evolution is that the two control parameters, mutation factor F and crossover rate CR are not kept fixed. Instead, each member of the population has individual values which are allowed to change between given ranges. When a new individual is formed, the offspring inherits the values from its progenitor, or new random values are generated with probability of τ_1 for F and τ_2 for CR. In this work the population size $NP = 10n_{dim}$ was used where n_{dim} is the number of dimensions in the decomposed design vector, i.e. each instance of algorithm uses an equal number of individuals in order to make comparing them fair. Mutation factor is allowed to vary within the range $F = [0.1, 1.0]$ and crossover rate $CR = [0.0, 1.0]$. The control parameter replacement probabilities are set to $\tau_{1,2} = 0.1$. The algorithms end when the stopping criteria $JS_k = 10^{-5}$ is reached.

Because the algorithms work in parallel, a generational approach would cause bottlenecks because of the non-constant fitness function computation times. Instead, a non-generational approach is used where the older individuals are replaced immediately if the offspring is superior. In addition, the elite information exchange is done asynchronously.

Three different approaches are tested. In the first one, the jDE algorithm is run traditionally using full domain and design vector. For the second approach, a "geometry decomposition" approach introduced in [9] is used ("Nash-jDE"). The design vector $\mathbf{x} = (r_1, \theta_1, \ldots, r_N, \theta_N)$ is divided between the elements $(\mathbf{x}_k = (r_k, \theta_k), k = \{1, \ldots, N\})$, which are then optimized using several jDE algorithms operating on separate subpopulations. After each generation, the global design vector is updated using elite values from each subpopulation. The proposed GNGCA algorithm is used in the third case. For flow reconstruction, since the flow is subsonic, the additive Schwarz domain decomposition algorithm is sufficient. The overlapped regions of subdomains are made of one strip. The computational domain is divided radially so that each subdomain contains one element (Fig. 1).

The FreeFEM++ v3.18 software is used as the solver [8]. The flow is computed using finite element method with a preconditioned conjugate gradient algorithm.

Table 1. Performance of the algorithms. The symbol n_{sl} refers to the number of (shape player) slave processes, t is the wall-clock time in seconds and n_{it} to the number of objective function evaluations required by the algorithm in order to reach the target precision.

		jDE		Nash-jDE		GNGCA		speed-up	
case	n_{sl}	t	n_{it}	t	n_{it}	t	n_{it}	jDE	N-jDE
2 elements	2	1155.00s	815	390.83s	279	306.57s	514	3.77×	1.27×
	4	332.05s	474	210.97s	302	194.74s	652	1.70×	1.08×
	6	190.42s	412	132.62s	279	174.60s	888	1.09×	0.76×
6 elements	6	3632.85s	4387	971.17s	1175	171.61s	1894	21.17×	5.66×
	12	1742.23s	4226	333.90s	809	115.87s	2502	15.04×	2.88×
	18	1201.11s	4369	244.53s	880	114.08s	3743	10.53×	2.14×

Since the flow is nonlinear, Eq. 3 is solved iteratively until the threshold value of $\varepsilon_\rho = 10^{-10}$ for density is reached. The algorithms are run on a computer containing 64 Intel Xeon CPU cores clocked at 2.67 GHz.

The mesh is constructed using Triangle v1.6 Delaunay mesh generator [10]. Numerical noise is minimized using mesh regeneration with the Laplacian. In order to avoid inverse elements and maintain mesh quality, the mesh is regenerated over certain intervals ($\delta r_k = 0.1$, $\delta\theta_k = 10^\circ$). An example decomposed mesh is illustrated in Fig. 3. Computing one subdomain gives speed-ups ranging from 3.2× to 14.0×.

4 Results and Discussion

The elapsed wall-clock times and the number of objective function evaluations required by each of the algorithm are listed on Table 1. Convergence curves of the algorithms are shown in Fig. 2. Final mesh and reconstructed global pressure field are compared to the reference in Fig. 3.

The results demonstrate that the geometry decomposition method using virtual Nash games can be used to increase algorithmic efficiency in geometry reconstruction problems. The proposed global Nash game approach shows that reconstructing geometry and flow simultaneously the wall-clock time can be reduced dramatically, provided the difference in the size of global and decomposed domains is sufficiently large. In the case of six domains, the speed-up compared to the original method is massive, over 20×. The increase compared to the pure geometry decomposition approach is also notable, over 5×. If the algorithms are compared a bit more fairly, i.e. the flow players are considered equal to the shape players, the speed-ups are 10× and 2×.

The efficiency of flow reconstruction is critical for the success of the proposed algorithm. Finding the correct geometry in an incompletely reconstructed flow field is not possible, which is evident in the large number of shape player objective function iterations needed. Unlike in the case of the other methods, increasing the number of slave processes brought only limited speed-ups for GNGCA. This was due the fact

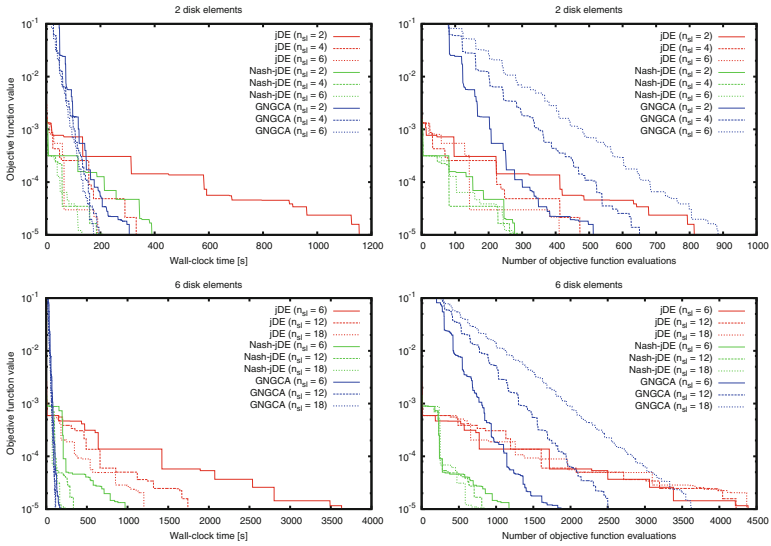


Fig. 2. Convergence curves of the tested algorithms. The onvergence according to the wall-clock time spent is on the *left* and the algorithmic convergence based on the required number of iterations is on the *right*

the flow players did not feed the shape players with accurate flow information fast enough resulting in an increased number of shape player iterations and correspondingly reduced efficiency improvement.

Algorithmic convergence can be improved by reducing the complexity of the problem. A classical method where the boundary nodes are used as shape design variables may be problematic due to a large number of variables. The situation can be improved using parallel algorithms and Bézier spline parametrization. In cases involving highly compressible potential flows where the flow is locally supersonic the domain reconstruction has to be augmented with an optimizer. The flow can be reconstructed using fast gradient methods on linearized equations coupled by DDM, or analogously to the shape presentation, the number of variables on interface boundary can be reduced using parametrization and the nonlinear flow can be reconstructed with evolutionary algorithms (cf. [3]).

5 Conclusion and Future

In this paper first results for a new "distributed one-shot" method that applies virtual Nash games, domain and geometry decomposition methods, are presented and discussed. The feasibility of the method is validated using an academic test case consisting of position reconstruction in a subsonic nonlinear flow.

In the forthcoming step, the Schwarz domain decomposition algorithm will be replaced with more robust methods. The simple compressible potential flow equa-

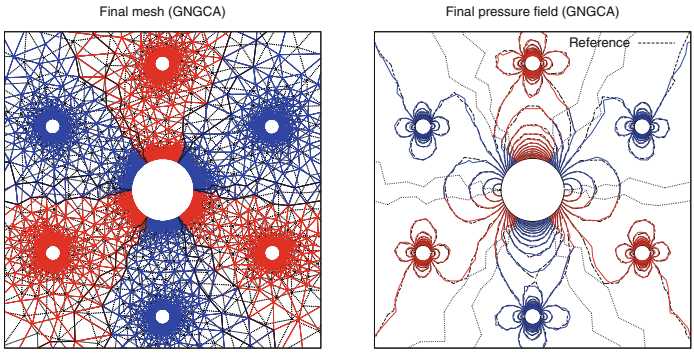


Fig. 3. Example final mesh and pressure field (GNGCA) compared to the reference

tion will be replaced with nonlinear systems of equations including Euler, Navier–Stokes, and Maxwell equations. Further tests involve complex geometries such as multi-element airfoils. The implementation of GPUs is also being studied. The ultimate target is to extend the method to speed up the capture of solutions of complex large scale problems which are frequently met in particular in 3D industrial detailed design.

Bibliography

[1] E. Arian and S. Ta'asan. Shape optimization in one shot. In *Optimal design and control*, pages 23–40. Birkhäuser Boston, Boston, MA, 1995.

[2] J. Brest, S. Greiner, B. Bošković, M. Mernik, and V. Žumer. Self-adapting control parameters in differential evolution: A comparative study on numerical benchmark problems. *IEEE Transactions on Evolutionary Computation*, 10(6): 646–657, December 2006.

[3] H.-Q. Chen, R. Glowinski, and J. Périaux. A domain decomposition/Nash equilibrium methodology for the solution of direct and inverse problems in fluid dynamics with evolutionary algorithms. In U. Langer et al., editor, *Domain Decomposition Methods in Science and Engineering XVIII*, Heidelberg, 2006. Springer.

[4] A. Jameson. Aerodynamic design via control theory. In *Recent advances in computational fluid dynamics*, pages 377–401. Springer, Berlin, 1989.

[5] J. F. Nash, Jr. Equilibrium points in *n*-person games. *Proc. Nat. Acad. Sci. U. S. A.*, 36:48–49, 1950. ISSN 0027–8424.

[6] E. Özkaya and N. R. Gauger. Single-step one-shot aerodynamic shape optimization. In *Optimal control of coupled systems of partial differential equations*, pages 191–204. Birkhäuser Verlag, Basel, 2009.

[7] J. Périaux, H.Q. Chen, B. Mantel, M. Sefrioui, and H.T. Sui. Combining game theory and genetic algorithms with application to DDM-nozzle optimization problems. *Finite Elem. Anal. Des.*, 37:417–429, 2001.

[8] O. Pironneau, F. Hecht, Le Hyaric A., and J. Morice. FreeFem++, www.freefem.org, 2012.

[9] M. Sefrioui and J. Périaux. Nash Genetic Algorithms: examples and applications. In *CEC00: Proceedings of the 2000 Congress on Evolutionary Computation*, pages 509–516. IEEE, November 2000.

[10] J. R. Shewchuk. Triangle: Engineering a 2D quality mesh generator and Delaunay triangulator. In M. C. Lin and D. Manocha, editors, *Applied Computational Geometry: Towards Geometric Engineering*, pages 203–222, Berlin, May 1996. Springer-Verlag.

[11] Z. Tang, J.-A. Désidéri, and J. Périaux. Distributed optimization using virtual and real game strategies for aerodynamic design. Technical Report 4543, INRIA, September 2002.

A Neumann-Dirichlet Preconditioner for FETI-DP Method for Mortar Discretization of a Fourth Order Problems in 2D

Leszek Marcinkowski*

Faculty of Mathematics, University of Warsaw, Banacha 2, 02-097 Warszawa, Poland,
Leszek.Marcinkowski@mimuw.edu.pl

1 Introduction

This study focuses on a construction of a parallel preconditioner for a FETI-DP (dual primal Finite Element Tearing and Interconnecting) method for a mortar Hsieh-Clough-Tocher (HCT) discretization of a model fourth order problem with discontinuous coefficients.

FETI-DP methods were introduced in [8]. They form a class of fast and efficient iterative solvers for algebraic systems of equations arising from the finite element discretizations of elliptic partial differential equations of second and fourth order, cf. [8, 10, 11, 16] and references therein. In a one-level FETI-DP method one has to solve a linear system for a set of dual variables formulated by eliminating all primal unknowns. The FETI-DP system contains in itself a coarse problem, while the preconditioner is usually fully parallel and constructed only from local problems.

There are many works investigating iterative solvers for mortar method for second order problem, e.g. cf. [1–3] and references therein. There have also been a few FETI-DP type algorithms developed for mortar discretization of second order problems, cf. e.g. [6, 7, 9]. But there is only a small number of studies focused on fast solvers for mortar discretizations of fourth order elliptic problems, cf. [12, 15, 17]. In this study we follow the approach of [9] which considers the case of a FETI-DP method for mortar discretization of a second order problem.

In this paper we first present the construction of mortar discretization of a fourth order elliptic problem which locally utilizes Hsieh-Clough-Tocher finite elements in the subdomains. Next we introduce a FETI-DP problem and then a Neumann-Dirichlet parallel preconditioner for a FETI-DP problem is proposed. Finally, we present the almost optimal bounds of the condition number, namely, a bound which grows like $C(1+\log(H/\underline{h}))^2$, where H is the maximal diameter of subdomains and $\underline{h}$ is a fine mesh parameter.

* This work was partially supported by Polish Scientific Grant 2011/01/B/ST1/01179.

R. Bank et al. (eds.), *Domain Decomposition Methods in Science and Engineering XX*, Lecture Notes in Computational Science and Engineering 91,
DOI 10.1007/978-3-642-35275-1_73,

2 Discrete Problem

In this section we focus on a mortar Hsieh-Clough-Tocher (HCT) finite element discretization for a model fourth order elliptic problem with discontinuous coefficients.

Let Ω be a polygonal domain in the plane. We assume that there exists a partition of Ω into disjoint polygonal subdomains Ω_k such that $\overline{\Omega} = \bigcup_{k=1}^{N} \overline{\Omega}_k$ with $\overline{\Omega}_k \cap \overline{\Omega}_l$ being an empty set, an edge or a vertex (crosspoint). We also assume that these subdomains form a coarse triangulation of the domain which is shape regular in the sense of [5]. We introduce a global interface $\Gamma = \bigcup_i \overline{\partial\Omega_i \setminus \partial\Omega}$ which plays an important role in our study.

Our model differential problem is to find $u^* \in H_0^2(\Omega)$ such that

$$a(u^*, v) = \int_\Omega f v \, dx \quad \forall v \in H_0^2(\Omega), \tag{1}$$

where $f \in L^2(\Omega)$, $H_0^2(\Omega) = \{u \in H^2(\Omega) : u = \partial_n u = 0 \quad \text{on} \quad \partial\Omega\}$ and $a(u,v) = \sum_{k=1}^{N} \int_{\Omega_k} \rho_k [u_{x_1x_1} v_{x_1x_1} + 2\, u_{x_1x_2} v_{x_1x_2} + u_{x_2x_2} v_{x_2x_2}] \, dx$. The coefficients ρ_k are positive and constant. Here $u_{x_k x_l} := \frac{\partial^2 u}{\partial x_k \partial x_l}$ for $k,l = 1,2$ and $\partial_n u$ is a unit normal derivative of u.

In each subdomain Ω_k we introduce a quasiuniform triangulation $T_h(\Omega_k)$ made of triangles with the parameter $h_k = \max_{\tau \in T_h(\Omega_k)} \operatorname{diam}(\tau)$, cf. e.g. [4]. We can now

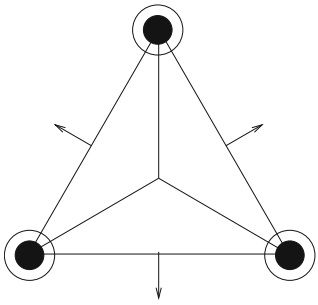


Fig. 1. Degrees of freedom of HCT element

introduce local finite element spaces. Let $X_h(\Omega_k)$ be the Hsieh-Clough-Tocher (HCT) macro finite element space defined as follows:

$$\begin{aligned} X_h(\Omega_k) = \{u \in C^1(\Omega_k) : {} & u \in P_3(\tau_i),\ \tau_i \in T_h(\Omega_k), \text{ for the subtriangles } \tau_i, \\ & i = 1,2,3, \text{ formed by connecting the vertices of} \\ & \text{any } \tau \in T_h(\Omega_k) \text{ to its centroid, and} \\ & u = \partial_n u = 0 \text{ on } \partial\Omega_k \cap \partial\Omega\}, \end{aligned}$$

where $P_3(\tau_i)$ is the function space of cubic polynomials defined over τ_i. The degrees of freedom of a function $u \in X_h(\Omega_k)$ over $\tau \in T_h(\Omega_k)$ are defined as: $\{u(p_k), \nabla u(p_k), \partial_n u(m_j)\}_{k,j=1,2,3}$, where p_k is a vertex and m_j is a midpoint of an edge of τ, cf. Fig. 1.

Next a global space $X_h(\Omega)$ is defined as $X_h(\Omega) = \prod_{i=1}^{N} X_h(\Omega_k)$. We also introduce $\widetilde{X}_h(\Omega)$ – a subspace of $X_h(\Omega)$ formed by all functions in $X_h(\Omega)$, which has all degrees of freedom continuous at the crosspoints, i.e. the common vertices of substructures.

Let Γ_{kl} denote the interface between two subdomains Ω_k and Ω_l i.e. the open edge that is common to these subdomains. Note that each interface Γ_{kl} inherits two one dimensional triangulations made of segments that are edges of elements of $T_h(\Omega_k)$ and $T_h(\Omega_l)$, respectively. Thus there are two independent 1D triangulations on Γ_{kl}: $T_{h,k}(\Gamma_{kl})$ related to Ω_k and another one associated with Ω_l - $T_{h,l}(\Gamma_{lk})$, cf. Fig. 2. Let γ_{kl} be a mortar, i.e. the side corresponding to Ω_k if $\rho_k \geq \rho_l$ and then let δ_{lk} be the other side of Γ_{lk} associated to Ω_l called a slave (nonmortar).

For each interface Γ_{kl} we introduce two test spaces associated with its slave triangulation $T_{h,l}(\delta_{lk})$ (cf. [13, 14]): let $M_t^h(\delta_{lk})$ be the space formed by C^1 smooth piecewise cubic functions on the slave triangulation of δ_{lk}, which are piecewise linear in the two end elements, and let $M_n^h(\delta_{lk})$ be the space of continuous piecewise quadratic functions on the elements of this triangulation, which are piecewise linear in the two end elements.

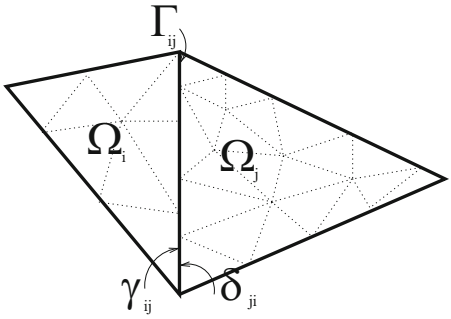


Fig. 2. Independent meshes on an interface Γ_{ij}

We also define a space $M = \prod_{\delta_{lk} \subset \Gamma} M_{lk}$ with $M_{lk} = M_t^l(\delta_{lk}) \times M_n^l(\delta_{lk})$ and a bilinear form $b(u, \psi)$: let $u = (u_k)_{k=1}^N \in \widetilde{X}_h(\Omega)$ and $\psi = (\psi_{lk})_{\delta_{lk}} = (\psi_{lk,t}, \psi_{lk,n})_{\delta_{lk}} \in M$, then $b(u, \psi) = \sum_{\delta_{lk} \subset \Gamma} \sum_{s \in \{t,n\}} b_{lk,s}(u, \psi_{lk,s})$ with

$$b_{lk,t}(u, \psi_{lk,t}) = \int_{\delta_{lk}} (u_k - u_l)\psi_{lk,t}\, ds,$$

$$b_{lk,n}(u, \psi_{lk,n}) = \int_{\delta_{lk}} (\partial_n u_k - \partial_n u_l)\psi_{lk,n}\, ds.$$

Further we will use the same notation for a function and for the vector with the values of degrees of freedom of this function.

We introduce discrete problem as the saddle point problem: find a pair $(u_h^*, \lambda^*) \in \widetilde{X}_h(\Omega) \times M$ such that

$$a(u_h^*, v) + b(v, \lambda^*) = f(v) \quad \forall v \in \widetilde{X}_h(\Omega), \tag{2}$$

$$b(u_h^*, \phi) = 0 \qquad \forall \phi \in M, \tag{3}$$

where $a_h(u,v) = \sum_{k=1}^N a_k(u,v)$ for

$$a_k(u,v) = \int_{\Omega_k} \rho_k [u_{x_1x_1} v_{x_1x_1} + 2\, u_{x_1x_2} v_{x_1x_2} + u_{x_2x_2} v_{x_2x_2}]\, dx.$$

This problem has a unique solution and error bounds are established, e.g. cf. [14].

3 Matrix Form of Mortar Conditions

Note that (3) is equivalent to two mortar conditions on each $\delta_{lk} = \gamma_{kl} = \Gamma_{kl}$:

$$\int_{\delta_{lk}} (u_k - u_l)\phi\, ds = 0 \quad \forall \phi \in M_t^l(\delta_{lk}), \tag{4}$$

$$\int_{\delta_{lk}} (\partial_n u_k - \partial_n u_l)\psi\, ds = 0 \quad \forall \psi \in M_n^l(\delta_{lk}). \tag{5}$$

We introduce the following splitting of two vectors representing the tangential and normal traces $u_{\delta_{lk}}$ and $\partial_n u_{\delta_{lk}}$: $u_{\delta_{lk}} = u^{(r)}_{\delta_{lk}} + u^{(c)}_{\delta_{lk}}$ and $\partial_n u_{\delta_{lk}} = \partial_n u^{(r)}_{\delta_{lk}} + \partial_n u^{(c)}_{\delta_{lk}}$ on a slave $\delta_{lk} \subset \partial\Omega_l$, where superscript (c) refers to degrees of freedom related to crosspoints (ends of this edge) and superscript (r) refers to degrees of freedom related to remaining nodes (vertices and midpoints) on this edge. We can now rewrite (4) and (5) in a matrix form on each interface $\Gamma_{kl} \subset \Gamma$:

$$\begin{aligned} B^{(c)}_{t,\delta_{lk}} u^{(c)}_{\delta_{lk}} + B^{(r)}_{t,\delta_{lk}} u^{(r)}_{\delta_{lk}} &= B^{(c)}_{t,\gamma_{kl}} u^{(c)}_{\gamma_{kl}} + B^{(r)}_{t,\gamma_{kl}} u^{(r)}_{\gamma_{kl}}, \\ B^{(c)}_{n,\delta_{lk}} \partial_n u^{(c)}_{\delta_{lk}} + B^{(r)}_{n,\delta_{lk}} \partial_n u^{(r)}_{\delta_{lk}} &= B^{(c)}_{n,\gamma_{kl}} \partial_n u^{(c)}_{\gamma_{kl}} + B^{(r)}_{n,\gamma_{kl}} \partial_n u^{(r)}_{\gamma_{kl}}, \end{aligned} \tag{6}$$

where the matrices $B_{t,\delta_{lk}} = [B^{(c)}_{t,\delta_{lk}},\, B^{(r)}_{t,\delta_{lk}}]$ and $B_{n,\delta_{lk}} = [B^{(c)}_{n,\delta_{lk}}, B^{(r)}_{n,\delta_{lk}}]$ are mass matrices obtained by substituting the traces of standard nodal basis functions of $X_h(\Omega_l)$ and nodal basis functions of $M_t^l(\delta_{lk}), M_n^l(\delta_{lk})$, respectively, into (4). The matrices $B_{t,\gamma_{kl}} = [B^{(c)}_{t,\gamma_{kl}},\, B^{(r)}_{t,\gamma_{kl}}]$ and $B_{n,\gamma_{kl}} = [B^{(c)}_{n,\gamma_{kl}}, B^{(r)}_{n,\gamma_{kl}}]$ are constructed analogously but utilizing traces onto γ_{kl} of standard nodal basis functions of $X_h(\Omega_k)$. Note that $B^{(r)}_{t,\delta_{lk}}, B^{(r)}_{n,\delta_{lk}}$ are positive definite square matrices, but that all other matrices in (6) are rectangular in general.

4 FETI-DP Problem

Let K_l be a matrix of $a_l(\cdot,\cdot)$ in the standard basis of $X_h(\Omega_l)$. Then let $\tilde{K}$ be the matrix obtained from a block diagonal matrix $K := \mathrm{diag}(K_l)_{l=1}^N$ by taking into account the continuity of the degrees of freedom at crosspoints. We can partition $\tilde{K}$ into

$$\tilde{K} = \begin{pmatrix} K_{ii} & K_{ic} & K_{ir} \\ K_{ci} & K_{cc} & K_{cr} \\ K_{ri} & K_{rc} & K_{rr} \end{pmatrix},$$

where the superscript (i) refer to the degrees of freedom associated with nodal points interior to subdomain, (c) to the degrees of freedom related to crosspoints, and (r) to the degrees of freedom associated the remaining nodes on masters and slaves. Then the matrix formulation of (2) and (3) is the following:

$$\begin{pmatrix} K_{ii} & K_{ic} & K_{ir} & 0 \\ K_{ci} & K_{cc} & K_{cr} & (B^{(c)})^T \\ K_{ri} & K_{rc} & K_{rr} & (B^{(r)})^T \\ 0 & B^{(c)} & B^{(r)} & 0 \end{pmatrix} \begin{pmatrix} u^{(i)} \\ u^{(c)} \\ u^{(r)} \\ \lambda^* \end{pmatrix} = \begin{pmatrix} f_i \\ f_c \\ f_r \\ 0 \end{pmatrix}. \tag{7}$$

Here $B^{(c)}$ is the matrix built from $B^{(c)}_{t,\delta_{lk}}, B^{(c)}_{n,\delta_{lk}}, B^{(c)}_{t,\gamma_{kl}}, B^{(c)}_{n,\gamma_{kl}}$ for all $\Gamma_{kl} = \gamma_{kl} = \delta_{lk} \subset \Gamma$ and $B^{(r)} := \mathrm{diag}([-B^{(r)}_{\gamma_{kl}}, B^{(r)}_{\delta_{lk}}])_{\Gamma_{kl} \subset \Gamma}$ is the block diagonal matrix with

$$B^{(r)}_{\gamma_{kl}} := \begin{pmatrix} B^{(r)}_{t,\gamma_{kl}} & 0 \\ 0 & B^{(r)}_{n,\gamma_{kl}} \end{pmatrix}, \qquad B^{(r)}_{\delta_{lk}} := \begin{pmatrix} B^{(r)}_{t,\delta_{lk}} & 0 \\ 0 & B^{(r)}_{n,\delta_{lk}} \end{pmatrix}. \tag{8}$$

Next we eliminate the unknowns related to the interior nodes and crosspoints i.e. $u^{(i)}$, $u^{(c)}$ in (7) and we get

$$\begin{aligned} \tilde{S}u^{(r)} + \tilde{B}^T\lambda^* &= \tilde{f}_r, \\ \tilde{B}u^{(r)} + \tilde{S}_{cc}\lambda^* &= \tilde{f}_c, \end{aligned} \tag{9}$$

where the respective matrices are defined as follows:

$$\tilde{S} := K_{rr} - (K_{ri}\, K_{rc})(\tilde{K}^{(ic)})^{-1} \begin{pmatrix} K_{ir} \\ K_{cr} \end{pmatrix},$$

$$\tilde{B} := B^{(r)} - (0\, B^{(c)})(\tilde{K}^{(ic)})^{-1} \begin{pmatrix} K_{ir} \\ K_{cr} \end{pmatrix},$$

and $\tilde{S}_{cc} := -(0\ B^{(c)})(\tilde{K}^{(ic)})^{-1} \begin{pmatrix} 0 \\ (B^{(c)})^T \end{pmatrix}$ with the nonsingular matrix $\tilde{K}^{(ic)} := \begin{pmatrix} K_{ii} & K_{ic} \\ K_{ci} & K_{cc} \end{pmatrix}$.

Eliminating $u^{(r)}$ we obtain the following FETI-DP problem: find $\lambda^* \in M$ such that

$$F(\lambda^*) = d, \tag{10}$$

where $d := \tilde{f}_c - \tilde{B}\tilde{S}^{-1}\tilde{f}_r$ and $F := \tilde{S}_{cc} - \tilde{B}\tilde{S}^{-1}\tilde{B}^T$.

5 Parallel Preconditioner

Let $W_r = \{w^{(r)} : w \in \widetilde{X}_h(\Omega)\}$ i.e. W_r is the space of vectors representing all degrees of freedom of functions from $\widetilde{X}_h(\Omega)$ associated with nodes (vertices and midpoints) on Γ but are *not* associated with crosspoints.

We can decompose any vector $w^{(r)} \in W_r$ into vectors related to masters and slaves:

$$w^{(r)} = \left(w_\Gamma^{(r)}, w_\Delta^{(r)}\right)^T,$$

where $w_\Gamma^{(r)}$ is the vector with the values of degrees of freedom which are associated with the nodes on the masters and $w_\Delta^{(r)}$ is the vector with the values of degrees of freedom which are related to the nodes on the slaves. We then introduce $W_\Delta = \{w_\Delta^{(r)} : w^{(r)} \in W_r\}$ i.e. the space formed by vectors in W_r which have only entries related to the degrees of freedom which are associated with the nodes on the slaves. It is very important to note that

$$\dim M = \dim W_\Delta.$$

Let S_Δ be the matrix obtained by restricting $\tilde{S} : W_r \to W_r$ to W_Δ.

Note that this matrix is can be represented as a block diagonal matrix with nonsingular diagonal blocks $S_{k,\Delta}$, i.e.

$$S_\Delta := \mathrm{diag}(S_{k,\Delta})_k,$$

where the subscript k runs over all subdomains that have at least one edge on Γ as a slave. Naturally, we could also partitioned this matrix with respect to the slaves.

Define nonsingular block diagonal matrix $B_\Delta : W_\Delta \to W_\Delta$:

$$B_\Delta := \mathrm{diag}(B_{\delta_{lk}}^{(r)})_{\delta_{lk} \subset \Gamma},$$

where $B_{\delta_{lk}}^{(r)}$ are block diagonal matrices (with two nonsingular blocks) defined in (8).

Then we introduce our parallel preconditioner:

$$\mathscr{M}_{DN}^{-1} := B_\Delta^{-T} S_\Delta B_\Delta^{-1},$$

which is nonsingular, or equivalently its inverse: $\mathscr{M}_{DN} := B_\Delta S_\Delta^{-1} B_\Delta^T$. Note that S_Δ and thus $\mathscr{M}_{DN}$ are dependent on the discontinuous coefficients ρ_k.

6 Condition Number Bounds

The main result of this paper is the following theorem which yields the bound of the condition number of preconditioned problem:

Theorem 1. *It holds that*

$$\langle \mathscr{M}_{DN}\lambda, \lambda\rangle \le \langle F\lambda, \lambda\rangle \le C\left(1 + \log\left(\frac{H}{\underline{h}}\right)\right)^2 \langle \mathscr{M}_{DN}\lambda, \lambda\rangle \qquad \forall \lambda \in M,$$

where $H = \max_k h_k$, $\underline{h} = \min_k h_k$, *and* C *a positive constant independent of the coefficients, or the parameters* H_k *and* h_k. *Here* $\langle \cdot, \cdot \rangle$ *is the standard* l_2 *inner product.*

As a direct consequence of this theorem we see that the condition number of $\mathscr{M}_{DN}^{-1}F$ is bounded by $C\left(1+\log\left(\frac{H}{h}\right)\right)^2$.

The lower bound in the theorem is obtained by purely algebraic arguments. And we get the upper bound by using several technical results of which the most important one is the estimate of special trace norms of jumps of tangential and normal traces over an interface $\Gamma_{kl} \subset \Gamma$.

Bibliography

[1] Yves Achdou, Yvon Maday, and Olof B. Widlund. Iterative substructuring preconditioners for mortar element methods in two dimensions. *SIAM J. Numer. Anal.*, 36(2):551–580, 1999.

[2] Petter E. Bjørstad, Maksymilian Dryja, and Talal Rahman. Additive Schwarz methods for elliptic mortar finite element problems. *Numer. Math.*, 95(3):427–457, 2003.

[3] Dietrich Braess, Wolfgang Dahmen, and Christian Wieners. A multigrid algorithm for the mortar finite element method. *SIAM J. Numer. Anal.*, 37(1):48–69, 1999.

[4] Susanne C. Brenner and L. Ridgway Scott. *The mathematical theory of finite element methods*, volume 15 of *Texts in Applied Mathematics*. Springer, New York, third edition, 2008.

[5] Susanne C. Brenner and Li-Yeng Sung. Balancing domain decomposition for nonconforming plate elements. *Numer. Math.*, 83(1):25–52, 1999.

[6] Nina Dokeva, Maksymilian Dryja, and Wlodek Proskurowski. A FETI-DP preconditioner with a special scaling for mortar discretization of elliptic problems with discontinuous coefficients. *SIAM J. Numer. Anal.*, 44(1):283–299, 2006.

[7] Maksymilian Dryja and Olof B. Widlund. A FETI-DP method for a mortar discretization of elliptic problems. In *Recent developments in domain decomposition methods (Zürich, 2001)*, volume 23 of *Lect. Notes Comput. Sci. Eng.*, pages 41–52. Springer, Berlin, 2002.

[8] Charbel Farhat, Michael Lesoinne, and Kendall Pierson. A scalable dual-primal domain decomposition method. *Numer. Linear Algebra Appl.*, 7(7–8):687–714, 2000. Preconditioning techniques for large sparse matrix problems in industrial applications (Minneapolis, MN, 1999).

[9] Hyea Hyun Kim and Chang-Ock Lee. A preconditioner for the FETI-DP formulation with mortar methods in two dimensions. *SIAM J. Numer. Anal.*, 42 (5):2159–2175, 2005.

[10] Axel Klawonn, Olof B. Widlund, and Maksymilian Dryja. Dual-primal FETI methods for three-dimensional elliptic problems with heterogeneous coefficients. *SIAM J. Numer. Anal.*, 40(1):159–179, 2002.

[11] Jan Mandel, Radek Tezaur, and Charbel Farhat. A scalable substructuring method by Lagrange multipliers for plate bending problems. *SIAM J. Numer. Anal.*, 36(5):1370–1391, 1999.

[12] Leszek Marcinkowski. Domain decomposition methods for mortar finite element discretizations of plate problems. *SIAM J. Numer. Anal.*, 39(4):1097–1114, 2001.

[13] Leszek Marcinkowski. A mortar element method for some discretizations of a plate problem. *Numer. Math.*, 93(2):361–386, 2002.

[14] Leszek Marcinkowski. A mortar finite element method for fourth order problems in two dimensions with Lagrange multipliers. *SIAM J. Numer. Anal.*, 42 (5):1998–2019, 2005.

[15] Leszek Marcinkowski. An Additive Schwarz Method for mortar Morley finite element discretizations of 4th order elliptic problem in 2d. *Electron. Trans. Numer. Anal.*, 26:34–54, 2007.

[16] Andrea Toselli and Olof Widlund. *Domain decomposition methods—algorithms and theory*, volume 34 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin, 2005.

[17] Xuejun Xu, Likang Li, and Wenbin Chen. A multigrid method for the mortar-type Morley element approximation of a plate bending problem. *SIAM J. Numer. Anal.*, 39(5):1712–1731, 2001/02.

A DG Space–Time Domain Decomposition Method

Martin Neumüller and Olaf Steinbach

Institute of Computational Mathematics, TU Graz, Steyrergasse 30, 8010 Graz, Austria,
neumueller@tugraz.at, o.steinbach@tugraz.at

Summary. In this paper we present a hybrid domain decomposition approach for the parallel solution of linear systems arising from a discontinuous Galerkin (DG) finite element approximation of initial boundary value problems. This approach allows a general decomposition of the space–time cylinder into finite elements, and is therefore applicable for adaptive refinements in space and time.

1 A Space–Time DG Finite Element Method

As a model problem we consider the transient heat equation

$$\partial_t u(x,t) - \Delta u(x,t) = f(x,t) \quad \text{for } (x,t) \in Q := \Omega \times (0,T), \tag{1}$$

$$u(x,t) = 0 \quad \text{for } (x,t) \in \Sigma := \partial\Omega \times (0,T), \tag{2}$$

$$u(x,0) = u_0(x) \quad \text{for } (x,t) \in \Omega \times \{0\} \tag{3}$$

where $\Omega \subset \mathbb{R}^n, n = 1,2,3$, is a bounded Lipschitz domain, and $T > 0$. Let $\mathscr{T}_N$ be a decomposition of the space–time cylinder $Q = \Omega \times (0,T) \subset \mathbb{R}^{n+1}$ into simplices τ_k of mesh size h. For simplicity we assume that the space–time cylinder Q has a polygonal ($n = 1$), a polyhedral ($n = 2$), or a polychoral ($n = 3$) boundary ∂Q. With $\mathscr{I}_N$ we denote the set of all interfaces (interior facets) e between two neighboring elements τ_k and τ_ℓ. For an admissible decomposition the interior facets are edges ($n = 1$), triangles ($n = 2$), or tetrahedrons ($n = 3$).

With respect to an interior facet $e \in \mathscr{I}_N$ we define for a function v the jump

$$[v]_e(x,t) := v_{|\tau_k}(x,t) - v_{|\tau_\ell}(x,t) \quad \text{for all } (x,t) \in e,$$

the average

$$\langle v \rangle_e(x,t) := \frac{1}{2}\left[v_{|\tau_k}(x,t) + v_{|\tau_\ell}(x,t)\right] \quad \text{for all } (x,t) \in e,$$

and the upwind in time direction by

R. Bank et al. (eds.), *Domain Decomposition Methods in Science and Engineering XX*, Lecture Notes in Computational Science and Engineering 91,
DOI 10.1007/978-3-642-35275-1_74,

$$\{v\}_e^{\rm up}(x,t) := \begin{cases} v_{|\tau_k}(x,t) & \text{for } n_t \geq 0, \\ v_{|\tau_\ell}(x,t) & \text{for } n_t < 0 \end{cases} \quad \text{for all } (x,t) \in e,$$

where $\mathbf{n} = (\mathbf{n}_x, n_t)$ is the normal vector of the interior facet e.

For a decomposition $\mathscr{T}_N$ of the space–time cylinder Q we introduce the discrete function space of piecewise polynomials of order p

$$S_{h,0}^p(\mathscr{T}_N) := \left\{ v : v_{|\tau_k} \in \mathbb{P}_p(\tau_k) \text{ for all } \tau_k \in \mathscr{T}_N \text{ and } v_{|\Sigma} = 0 \right\}.$$

The proposed space–time approach is based on the use of an interior penalty Galerkin approximation of the Laplace operator and an upwind scheme for the approximation of the time derivative, see, e.g., [3, 5]. Hence we have to find $u_h \in S_{h,0}^p(\mathscr{T}_N)$ such that

$$\begin{aligned} a_{\rm DG}(u_h, v_h) &:= -\sum_{k=1}^N \int_{\tau_k} u_h \, \partial_t v_h \, dxdt + \int_{\Sigma_T} u_h \, v_h \, dx \\ &+ \sum_{e \in \mathscr{I}_N} \int_e n_t \, \{u_h\}_e^{\rm up} [v_h]_e \, ds_{(x,t)} + \sum_{k=1}^N \int_{\tau_k} \nabla_x u_h \cdot \nabla_x v_h \, dxdt \\ &- \sum_{e \in \mathscr{I}_N} \int_e \left[\langle \mathbf{n}_x \cdot \nabla_x u_h \rangle_e \, [v_h]_e - \varepsilon \, [u_h]_e \langle \mathbf{n}_x \cdot \nabla_x v_h \rangle_e \right] ds_{(x,t)} \\ &+ \frac{\sigma}{h} \sum_{e \in \mathscr{I}_N} \int_e |\mathbf{n}_x|^2 \, [u_h]_e \, [v_h]_e \, ds_{(x,t)} \\ &= \int_Q f \, v_h \, dxdt + \int_{\Sigma_0} u_0 \, v_h \, dx =: F(v_h) \end{aligned} \tag{4}$$

is satisfied for all $v_h \in S_{h,0}^p(\mathscr{T}_N)$. The parameters σ and ε have to be chosen appropriately. For $v_h \in S_{h,0}^p(\mathscr{T}_N)$ and $\sigma > 0$ the related energy norm is given by

$$\|v_h\|_{\rm DG}^2 := \|v_h\|_A^2 + \|v_h\|_B^2,$$

where

$$\begin{aligned} \|v_h\|_A^2 &:= \sum_{k=1}^N \|\nabla_x v_h\|_{\tau_k}^2 + \frac{\sigma}{h} \sum_{e \in \mathscr{I}_N} \| \, |\mathbf{n}_x| \, [v_h]_e \|_{L_2(e)}^2, \\ \|v_h\|_B^2 &:= h \sum_{k=1}^N \|\partial_t v_h\|_{\tau_k}^2 + \frac{1}{2} \|v_h\|_{L_2(\Sigma_0 \cup \Sigma_T)}^2 + \frac{1}{2} \sum_{e \in \mathscr{I}_N} \| \sqrt{|n_t|} \, [v_h]_e \|_{L_2(e)}^2. \end{aligned}$$

The unique solvability of the variational formulation (4) is based on the following stability result.

Lemma 1. *Let $\varepsilon \in \{-1, 0, 1\}$ and $\sigma > 0$. For $\varepsilon \in \{-1, 0\}$ let σ be sufficient large. Then the stability estimate*

$$\sup_{0 \neq v_h \in S_{h,0}^p(\mathscr{T}_N)} \frac{a_{\rm DG}(u_h, v_h)}{\|v_h\|_{\rm DG}} \geq c_1^A \|u_h\|_{\rm DG} \quad \textit{for all } u_h \in S_{h,0}^p(\mathscr{T}_N)$$

is satisfied where the constant c_1^A depends on the shape of the finite elements, and on the stabilization parameter σ. However, for a sufficient large choice of σ we can ensure $c_1^A = \frac{1}{2}$.

Proof. The proof follows as in [5], by using the technique as in [2]; see also [3]. □

By using standard arguments we can then conclude the energy error estimate

$$\|u - u_h\|_{\mathrm{DG}} \leq ch^{\min\{s,p+1\}-1}|u|_{H^s(Q)}$$

when assuming $u \in H^s(Q)$ for some $s \leq p+1$, and, by applying the Aubin–Nitsche trick, for $\varepsilon = -1$,

$$||u - u_h||_{L_2(\Omega)} \leq ch^{\min\{s,p+1\}}\,|u|_{H^s(Q)}\,. \tag{5}$$

To illustrate the proposed DG finite element method in space and time as well as the given error estimates we consider a first numerical example for the initial boundary value problem (1)–(3) for $n = 1$ and $\Omega = (0,1)$, $T = 1$. This implies $Q = (0,1)^2$. The given data f and u_0 are chosen such that the solution is given as

$$u(x,t) = \sin(\pi x)(1-t)^{3/4} \in H^{1.25-\bar{\varepsilon}}(Q) \quad \text{with } \bar{\varepsilon} > 0.$$

Starting from a triangulation of $Q = (0,1)^2$ into four triangles we consider a sequence of several uniform refinement steps to analyze the convergence behavior of the presented method. Using piecewise linear basis functions, i.e. $p = 1$, $\varepsilon = -1$ and $\sigma = 10$, the numerical results are given in Table 1 which confirm the convergence rate of 1.25 as predicted by the error estimate (5).

level	elements	dof	$\|\|u - u_h\|\|_{L_2(Q)}$	eoc
0	4	8	$2.2679-1$	–
1	16	40	$5.1354-2$	2.14
2	64	176	$1.3107-2$	1.97
3	256	736	$3.4813-3$	1.91
4	1024	3008	$9.7383-4$	1.84
5	4096	12160	$3.0406-4$	1.68
6	16384	48896	$1.0923-4$	1.48
7	65536	196096	$4.3315-5$	1.33
8	262144	785408	$1.7935-5$	1.27
9	1048576	3143680	$7.5278-6$	1.25
10	4194304	12578816	$3.1694-6$	1.25
11	16777216	50323456	$1.3345-6$	1.25

Table 1. Numerical results for $p = 1$, $\varepsilon = -1$ and $\sigma = 10$.

2 A Hybrid Space–Time Domain Decomposition Method

The presented space–time method (4) results in a large linear system of algebraic equations. For its iterative solution we introduce a hybrid formulation as in [1, 2]. Therefore we subdivide the space–time domain Q into P non–overlapping subdomains Q_i, $i = 1, \ldots, P$,

$$\overline{Q} = \bigcup_{i=1}^{P} \overline{Q}_i, \quad Q_i \cap Q_j = \emptyset \quad \text{for } i \neq j.$$

By

$$\Gamma := \bigcup_{i=1}^{P} \Gamma_i \quad \text{with } \Gamma_i := \overline{\partial Q_i \setminus \partial Q}$$

we denote the interface of the space–time domain decomposition, see Fig. 1.

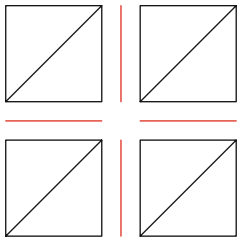


Fig. 1. Space–time decomposition of Q and the interface Γ

With respect to the interface Γ we introduce the discrete function space of piecewise polynomials of order p,

$$S_h^p(\Gamma) := \left\{ v \in L_2(\Gamma) : v_{|e} \in \mathbb{P}_p(e) \text{ for all } e \in \mathscr{I}_N \text{ with } e \subseteq \Gamma \right\}.$$

For the solution of the local partial differential equations in all subdomains Q_i we apply the space–time method as described by the variational formulation (4). For this we denote by $a_{\mathrm{DG}}^{(i)}(\cdot,\cdot)$ the restriction of the bilinear form $a_{\mathrm{DG}}(\cdot,\cdot)$ on the subdomain Q_i, $i = 1, \ldots, P$, i.e.

$$\begin{aligned}
a_{\mathrm{DG}}^{(i)}(u_h, v_h) := & - \sum_{k=1}^{N} \int_{\tau_k \cap Q_i} u_h \, \partial_t v_h \, dx dt + \int_{\Sigma_T \cap \partial Q_i} u_h \, v_h \, dx \\
& + \sum_{e \in \mathscr{I}_N} \int_{e \cap Q_i} n_t \, \{u_h\}_e^{\mathrm{up}} \, [v_h]_e \, ds_{(x,t)} + \sum_{k=1}^{N} \int_{\tau_k \cap Q_i} \nabla_x u_h \cdot \nabla_x v_h \, dx dt \\
& - \sum_{e \in \mathscr{I}_N} \int_{e \cap Q_i} \left[\langle \mathbf{n}_x \cdot \nabla_x u_h \rangle_e \, [v_h]_e - \varepsilon \, [u_h]_e \, \langle \mathbf{n}_x \cdot \nabla_x v_h \rangle_e \right] ds_{(x,t)} \\
& + \frac{\sigma}{h} \sum_{e \in \mathscr{I}_N} \int_{e \cap Q_i} |\mathbf{n}_x|^2 \, [u_h]_e [v_h]_e ds_{(x,t)}.
\end{aligned}$$

Accordingly, the restriction of the linear form $F(\cdot)$ on a subdomain Q_i is given by

$$F^{(i)}(v_h) := \int_{Q_i} f\, v_h \, dxdt + \int_{\Sigma_0 \cap \partial Q_i} u_0\, v_h \, dx.$$

For the coupling of the local fields we first introduce a new unknown on the interface,

$$\lambda := \langle u \rangle_e = \frac{1}{2}\left[u_{|\tau_k} + u_{|\tau_\ell}\right] \quad \text{on } \Gamma \cap e.$$

With this we can rewrite the jump of a function as

$$[u]_e = u_{|\tau_k} - u_{|\tau_\ell} = 2\left(u_{|\tau_k} - \lambda\right) = 2\left(\lambda - u_{|\tau_\ell}\right) \quad \text{on } \Gamma \cap e.$$

Therefore we obtain for the coupling terms related to the Laplace operator

$$\sum_{e \in \mathscr{I}_N} \int_{e \cap \Gamma} \langle \mathbf{n}_x \cdot \nabla_x u \rangle_e \, [v]_e \, ds_{(x,t)} = \sum_{k=1}^{N} \int_{\partial \tau_k \cap \Gamma} \mathbf{n}_{k,x} \cdot \nabla_x u \, (v - \mu) \, ds_{(x,t)},$$

$$\sum_{e \in \mathscr{I}_N} \int_{e \cap \Gamma} [u]_e \, \langle \mathbf{n}_x \cdot \nabla_x v \rangle_e \, ds_{(x,t)} = \sum_{k=1}^{N} \int_{\partial \tau_k \cap \Gamma} (u - \lambda) \, \mathbf{n}_{k,x} \cdot \nabla_x v \, ds_{(x,t)},$$

$$\sum_{e \in \mathscr{I}_N} \int_{e \cap \Gamma} |\mathbf{n}_x|^2 \, [u]_e \, [v]_e \, ds_{(x,t)} = 2 \sum_{k=1}^{N} \int_{\partial \tau_k \cap \Gamma} |\mathbf{n}_{k,x}|^2 \, (u - \lambda)(v - \mu) \, ds_{(x,t)}.$$

For the classical solution u of the transient heat equation (1)–(3) there obviously holds for an interior facet $e \in \mathscr{I}_N$

$$\lambda = \langle u \rangle_e = \frac{1}{2}\left[u_{|\tau_k} + u_{|\tau_\ell}\right] = u_{|\tau_k} = u_{|\tau_\ell} \quad \text{on } e.$$

Therefore the upwind in time can be written as

$$\{u\}_e^{\mathrm{up}} = \begin{cases} u_{|\tau_k} & \text{for } n_t \geq 0, \\ u_{|\tau_\ell} & \text{for } n_t < 0 \end{cases} = \begin{cases} u_{|\tau_k} & \text{for } n_{k,t} \geq 0, \\ \lambda & \text{for } n_{k,t} < 0 \end{cases} =: \{u/\lambda\}_{\partial \tau_k}^{\mathrm{up}} \quad \text{on } \Gamma \cap e.$$

The coupling containing the upwind part can now be expressed by

$$\sum_{e \in \mathscr{I}_N} \int_{e \cap \Gamma} n_t \, \{u\}_e^{\mathrm{up}} \, [v]_e \, ds_{(x,t)} = \sum_{k=1}^{N} \int_{\partial \tau_k \cap \Gamma} n_{k,t} \, \{u/\lambda\}_{\partial \tau_k}^{\mathrm{up}} (v - \mu) \, ds_{(x,t)}.$$

With respect to each subdomain Q_i we therefore can define the bilinear form

$$\begin{aligned} c^{(i)}(u_h, \lambda_h; v_h, \mu_h) := & \sum_{\substack{k=1 \\ \tau_k \subseteq Q_i}}^{N} \int_{\partial \tau_k \cap \Gamma} n_{k,t} \, \{u_h/\lambda_h\}_{\partial \tau_k}^{\mathrm{up}} (v_h - \mu_h) \, ds_{(x,t)} \\ & - \sum_{\substack{k=1 \\ \tau_k \subseteq Q_i}}^{N} \int_{\partial \tau_k \cap \Gamma} \left[\mathbf{n}_{k,x} \cdot \nabla_x u_h \, (v_h - \mu_h) - \varepsilon (u_h - \lambda_h) \, \mathbf{n}_{k,x} \cdot \nabla_x v_h\right] ds_{(x,t)} \\ & + \frac{2\sigma}{h} \sum_{\substack{k=1 \\ \tau_k \subseteq Q_i}}^{N} \int_{\partial \tau_k \cap \Gamma} |\mathbf{n}_{k,x}|^2 \, (u_h - \lambda_h)(v_h - \mu_h) \, ds_{(x,t)}. \end{aligned}$$

Hence we can write the discrete hybrid space–time variational formulation to find $u_h \in S^p_{h,0}(\mathscr{T}_N)$ and $\lambda_h \in S^p_h(\Gamma)$ satisfying

$$\sum_{i=1}^{P} \left[a^{(i)}_{\mathrm{DG}}(u_h, v_h) + c^{(i)}(u_h, \lambda_h; v_h, \mu_h) \right] = \sum_{i=1}^{P} F^{(i)}(v_h) \tag{6}$$

for all $v_h \in S^p_{h,0}(\mathscr{T}_N)$ and $\mu_h \in S^p_h(\Gamma)$. As in [2] we can prove unique solvability of the hybrid scheme (6). Moreover, related error estimates as derived for the DG scheme remain valid.

The discrete variational formulation (6) is equivalent to the solution of the linear equations

$$\begin{pmatrix} A^{(1)}_{II} & & & A^{(1)}_{I\Gamma} \\ & A^{(2)}_{II} & & A^{(2)}_{I\Gamma} \\ & & \ddots & \vdots \\ & & A^{(P)}_{II} & A^{(P)}_{I\Gamma} \\ A^{(1)}_{\Gamma I} & A^{(2)}_{\Gamma I} & \cdots & A^{(P)}_{\Gamma I} \; A_{\Gamma\Gamma} \end{pmatrix} \begin{pmatrix} \mathbf{u}^{(1)}_I \\ \mathbf{u}^{(2)}_I \\ \vdots \\ \mathbf{u}^{(P)}_I \\ \lambda_\Gamma \end{pmatrix} = \begin{pmatrix} \mathbf{f}^{(1)}_I \\ \mathbf{f}^{(2)}_I \\ \vdots \\ \mathbf{f}^{(P)}_I \\ \mathbf{f}_\Gamma \end{pmatrix} \tag{7}$$

where the local block matrices $A^{(i)}_{II}$ correspond to the local bilinear forms $a^{(i)}_{\mathrm{DG}}(\cdot,\cdot)$ and $c^{(i)}(\cdot,0;\cdot,0)$, while the remaining block matrices describe the coupling across the interface. For an appropriate choice of the DG parameters, see Lemma 1, the local matrices $A^{(i)}_{II}$ are invertible. Hence we obtain the Schur complement system

$$\left[A_{\Gamma\Gamma} - \sum_{i=1}^{P} A^{(i)}_{\Gamma I} \left(A^{(i)}_{II} \right)^{-1} A^{(i)}_{I\Gamma} \right] \lambda_\Gamma = \mathbf{f}_\Gamma - \sum_{i=1}^{P} A^{(i)}_{\Gamma I} \left(A^{(i)}_{II} \right)^{-1} \mathbf{f}^{(i)}_I, \tag{8}$$

with

$$\mathbf{u}^{(i)}_I = \left(A^{(i)}_{II} \right)^{-1} \left[\mathbf{f}^{(i)}_I - A^{(i)}_{I\Gamma} \lambda_\Gamma \right] \quad \text{for } i = 1, \ldots, P.$$

The inversion of the local matrices $A^{(i)}_{II}$ can be done in parallel either by using some appropriate direct approach, or suitable iterative schemes. For the solution of the global Schur complement system (8) we can use, for example the GMRES method.

3 Numerical Examples

To illustrate the hybrid domain decomposition approach we consider for $n = 3$ the spatial domain $\Omega = (0,1)^3$ and $T = 1$, i.e. $Q = (0,1)^4$. As initial triangulation for the space–time domain we use 96 pentatopes of the same size, see also [4]. The initial triangulation is used as a partition of the space–time domain into $P = 96$ subdomains, which we keep fixed for all computations. As exact solution of the transient heat equation (1) we now consider the smooth function

$$u(x,t) = \sin(\pi x_1) \sin(\pi x_2) \sin(\pi x_3) t^2.$$

For the iterative solution of the Schur complement system (8) we use the GMRES method without preconditioning with a relative error reduction of $\varepsilon_{\text{GMRES}} = 10^{-8}$. In the Tables 2 and 3 we present the iteration numbers of the GMRES method for different levels of a uniform refinement of the space–time mesh for $p = 1$ and $p = 2$. We observe that the number of required iterations grows slightly indicating the need of using an appropriate preconditioner. The results also show the optimal convergence rates for the error in the $L_2(Q)$ norm when using linear and quadratic basis functions.

level	elements	dof $\mathbf{u}_I^{(i)}$	dof λ_Γ	iter.	$\|\|u-u_h\|\|_{L_2(Q)}$	eoc
0	96	192	768	68	$6.120-2$	–
1	1536	5376	6144	143	$3.821-2$	0.68
2	24576	104448	49152	197	$1.356-2$	1.49
3	393216	1818624	393216	294	$4.024-3$	1.75
4	6291456	30277632	3145728	475	$1.111-3$	1.86

Table 2. Numerical results with 96 subdomains for $p = 1$, $\varepsilon = -1$ and $\sigma = 10$.

level	elements	dof $\mathbf{u}_I^{(i)}$	dof λ_Γ	iter.	$\|\|u-u_h\|\|_{L_2(Q)}$	eoc
0	96	720	1920	404	$4.199-2$	–
1	1536	17280	15360	699	$7.492-3$	2.49
2	24576	322560	122880	900	$1.005-3$	2.90
3	393216	5529600	983040	1131	$1.293-4$	2.96

Table 3. Numerical results with 96 subdomains for $p = 2$, $\varepsilon = -1$ and $\sigma = 10$.

4 Conclusions

In this paper we have presented a hybrid DG domain decomposition approach for the parallel solution of initial boundary value problems. Numerical examples for one– and three–dimensional spatial domains indicate the accuracy and applicability of the proposed method. However, the numerical results also indicate the need to use an appropriate global preconditioner for the Schur complement system (8). Moreover, when solving the coupled system (7) iteratively, suitable local preconditioners are mandatory as well. A possible choice is to use space–time multigrid methods. Although we have only considered uniform refinements in this paper, the proposed approach is also applicable to non–uniform and adaptive refinements, see, for example, [4]. For this we need to use suitable a posteriori error estimators, and the solution algorithms need to be robust with respect to adaptive refinements. Although we have

only considered the simple model problem of the transient heat equation, the proposed approach can be extended to more complicated problems, see, e.g., [4] for a first example for the transient Navier–Stokes system.

Acknowledgments This work was supported by the Austrian Science Fund (FWF) within the SFB Mathematical Optimization and Applications in Biomedical Sciences.

Bibliography

[1] B. Cockburn, J. Gopalakrishnan, and R. Lazarov. Unified hybridization of discontinuous Galerkin, mixed, and continuous Galerkin methods for second order elliptic problems. *SIAM J. Numer. Anal.*, 47:1319–1365, 2009.

[2] H. Egger and J. Schöberl. A hybrid mixed discontinuous Galerkin finite–element method for convection–diffusion problems. *IMA J. Numer. Anal.*, 30:1206–1234, 2010.

[3] M. Neumüller. Eine Finite Element Methode für optimale Kontrollprobleme mit parabolischen Randwertaufgaben. Masterarbeit, Institut für Numerische Mathematik, Technische Universität Graz, 2010.

[4] M. Neumüller and O. Steinbach. Refinement of flexible space–time finite element meshes and discontinuous Galerkin methods. *Comput. Visual. Sci.*, 14:189–205, 2011.

[5] B. Rivière. *Discontinuous Galerkin Methods for Solving Elliptic and Parabolic Equations*. SIAM, Philadelphia, 2008.

Parallel Adaptive Deflated GMRES

Désiré Nuentsa Wakam[1], Jocelyne Erhel[1], and William D. Gropp[2]

[1] INRIA Campus de Beaulieu 35042 Rennes Cedex, {desire.nuentsa_wakam, jocelyne.erhel}@inria.fr
[2] NCSA, University of Illinois Urbana-Champaign, wgropp@illinois.edu

Summary. Many scientific libraries are currently based on the GMRES method as a Krylov subspace iterative method for solving large linear systems. The restarted formulation known as GMRES(m) has been extensively studied and several approaches have been proposed to reduce the negative effects due to the restarting procedure. A common effect in GMRES(m) is a slow convergence rate or a stagnation in the iterative process. In this situation, it is less attractive as a general solver in industrial applications. In this work, we propose an adaptive deflation strategy which retains useful information at time of restart to avoid stagnation in GMRES(m) and improve its convergence rate. We give a parallel implementation in the PETSc package. The provided numerical results show that this approach can be effectively used in the hybrid direct/iterative methods to solve large-scale systems.

1 Introduction

The GMRES method due to [11] is widely used, thanks to its monotonic convergence properties, as a Krylov subspace method for solving large and sparse linear systems. Due to memory and computational requirements, the restarted GMRES (noted as GMRES(m)) is generally used. At the time of restart, information from the previous Krylov subspace is discarded and the orthogonality between successive Krylov subspaces is not preserved. The worst case is when the successive generated Krylov subspaces are very close. As a result, there is no significant reduction in the residual norm and the iterative process may stagnate. Deflation techniques are a class of acceleration strategies that collects useful information at the time of restart mainly to avoid this stagnation and improve the convergence rate. The main idea behind these methods is to remove the smallest eigencomponents from the residual vector as they are known to slow down the convergence of GMRES.

In a practical use of a deflation strategy, it is necessary to define the number of eigenvalues to deflate. As the deflation process induces additional operations to GMRES(m), it is interesting as well to know a priori if the deflation will be beneficial. In this work, we propose an adaptive deflated GMRES(m) which aims at enhancing the convergence of GMRES(m) by adaptively extracting the spectral information

R. Bank et al. (eds.), *Domain Decomposition Methods in Science and Engineering XX*, Lecture Notes in Computational Science and Engineering 91,
DOI 10.1007/978-3-642-35275-1_75,

needed to speedup the convergence. The adaptive strategy is based on a (near) stagnation test which defines if the deflation process is needed or not and if more accurate spectral information are required. Although we use a stagnation test similar to that in [12], our approach is different since we assume that the restart length m is fixed. This work is motivated by the convergence behavior of GMRES when it is used with a Schwarz preconditioner. As the number of subdomains increases, the eigenvalues are less and less clustered. The restarting may have the disadvantage to discard the smallest eigenvalues before their convergence. The proposed adaptive strategy will thus keep these spectral values in the Krylov subspace until their convergence.

The remaining part of this report is organized as follows: in Sect. 2, we first recall the basis of the deflation technique applied as a preconditioner and we derive the adaptive strategy. In Sect. 3, we discuss on the parallel implementation. Section 4 is focused on numerical experiments to show the benefits of this scheme on a real industrial CFD test case.

2 Adaptive Preconditioner for the Deflated GMRES(m)

We are interested in the solution of the linear system

$$Ax = b \tag{1}$$

The GMRES method is among the best methods to solve this system when the coefficient matrix A is nonsingular and nonsymmetric. For large linear systems, the restarted version should always be used to reduce the memory and computational requirements. The deflated GMRES has been proposed to reduce the negative effects of the restarting procedure. The general idea behind these methods is to add to the Krylov subspace an approximation of the invariant subspace associated to the smallest eigenvalues. In [7], this is carried out by defining a preconditioner that is equal to the projected matrix onto the approximated invariant subspace and is taken as the identity on the orthogonal subspace. Hence, given $U = [u_1, \ldots, u_r] \in \mathbb{R}^{n \times r}$ the r-dimensional basis of the invariant subspace associated to the eigenvalues to deflate, the preconditioner is defined as

$$M_D^{-1} \equiv I_n + U(|\lambda_n| T_r^{-1} - I_r)U^T, \quad T = U^T B U, \tag{2}$$

where λ_n is the largest eigenvalue in magnitude, I_n and I_r are the identity matrices and B the initial preconditioned matrix. Since M_D^{-1} is nonsingular, the eigenvalues of the resulted matrix $M_D^{-1}B$ or BM_D^{-1} are $\lambda_{r+1}, \ldots, \lambda_n, |\lambda_n|$ with a multiplicity at least r. It is therefore expected to get a faster convergence rate with this preconditioner since the r smallest eigencomponents that slow down the convergence are deflated. This assumes that U is a good approximation of the basis of the selected invariant subspace. For large matrices however, the cost of accurately computing U (as suggested in [7] and later in [4]) may induce a significant overhead. This process should be carried out only when it is necessary, for instance to avoid stagnation.

Algorithm 4 DGMRES(m,k,r): Restarted GMRES with adaptive deflation

input (m, $itmax$, ε, k, smv, bgv, $rmax$);
Set $B \equiv AM^{-1}$, M^{-1} is any external preconditioner
$r_0 = b - Ax_0$; $U = [\,]$; $M_D = I$; $it = 0$; $r = 0$;
while ($\|r_0\| > \varepsilon$)
  Arnoldi process on B to get $BM_D^{-1}V_m = V_{m+1}\bar{H}_m$. See [11]
  $x_m = x_0 + M_D^{-1}M^{-1}V_m y_m$, y_m solution of $min\|\beta e_1 - \bar{H}_m y_m\|_2$;
  $r_m = b - Ax_m$, $it \leftarrow it + m$;
  If ($\|r_m\| > \varepsilon$ and $it < itmax$) **then**
    $Iter = m * log(\frac{\varepsilon}{\|r_m\|})/log(\frac{\|r_m\|}{\|r_0\|})$;
    If($(Iter > smv * (itmax - it)$ **and** $r < rmax$) **then**
      Compute k Schur vectors of B noted X. See [7]
      Orthogonalize X against U
      Compute $T = \begin{bmatrix} U & X \end{bmatrix}^T B \begin{bmatrix} U & X \end{bmatrix} \equiv \begin{pmatrix} U^TBU & U^TBX \\ X^TBU & X^TBX \end{pmatrix}$
      Increase U by X; $r \leftarrow r + k$;
      If($Iter > bgv * (itmax - it)$) **then**
        Improve U as indicated in [4, Sect. 3]
      EndIf
      Factorize T Set $M_D^{-1} \equiv I_n + U(|\lambda_n|T^{-1} - I_r)U^T$
    End If
  End If
  $x_0 = x_m$, $r_0 = r_m$
end while

We thus propose here an adaptive strategy that detects a near-stagnation in the iterative process or a slow reduction in the residual norm. This approach is based upon the work by Sosonkina et al. [12] in which the Krylov subspace is adaptively increased along the cycles of GMRES(m); Here, we find it natural to enrich the subspace with the eigencomponents that slow down the convergence. The main steps are given in Algorithm 4. First, m steps of the Arnoldi process are performed to compute the orthonormal basis V_m. It also creates an upper Hessenberg matrix $H_m = V_m^T BV_m$ which is the restriction of B onto the $m-$dimensional Krylov subspace. Then, a least-squares problem is solved to minimize the residual norm in the Krylov subspace. At the time of restart, if the desired residual norm is not achieved, a stagnation test is computed to determine if a deflation process could be beneficial to accelerate the convergence. This test considers the convergence rate over the previous restart cycles and evaluates the number of iterations ($Iter$) needed to achieve the desired accuracy. If $Iter$ is greater than the remaining number of steps (bounded by a small multiple smv of the number of iterations allowed), then data are computed to update the preconditioner associated to the deflation process. This test is therefore used to reduce the iteration counts in GMRES(m). To detect a near-stagnation, we use another test which considers a large multiple bgv of the remaining number of steps. In this case, a harmonic projection is carried out to accurately compute the eigenvalues and continuously update the previous estimation of U.

3 Implementation Notes

We now give some details about the implementation of Algorithm 4 on distributed-memory computers. The programming model is SPMD (Single Program Multiple Data) and communications are done using the message-passing interface (MPI). The adjacency graph of the input sparse matrix is first built. PARMETIS is then used to partition the vertices of the graph into D disjoint vertices. From this partitioning, the matrix is distributed such that each processor holds a contiguous chunk of rows corresponding to the vertices it owns. The right hand side and all other vectors (Krylov basis, invariant basis) are distributed accordingly. Note that the goal of this data distribution is to get a good load balance and to minimize communication during matrix-vector multiply and preconditioning steps. When the additive Schwarz preconditioner is used, an overlapping partitioning can be defined by taking recursively adjacent vertices from the initial disjoint partitions.

The main parallel operations in Algorithm 4 so far are the matrix-vector multiply, scalar products, and the application of M^{-1} and M_D^{-1}. M^{-1} can be any parallel preconditioner as long as it implements the basic operation $v_j \leftarrow M^{-1}v_i$. In our tests, the restricted additive Schwarz has been used as defined in [5]. It is then necessary in the setup phase to factorize in each process the block matrices A_p corresponding to the restriction of A onto the defined subdomains. M_D^{-1} is applied to a distributed vector v_j in a straightforward manner given the data distribution described above. This implies r all-to-all communications to compute the projection onto the invariant subspace. There is no additional communication for the other terms since the $r \times r$ dense matrix T is owned by each process.

We provide an implementation of this method using the PETSc package (see [3]). The original implementation of the built-in *KSP GMRES* has been modified to provide the data needed for the deflation and to apply the resulting preconditioner to generate the Krylov basis. Although the current presentation does not discuss the choice of side of preconditioning, the implementation does define left and right preconditioning. Note that the current adaptive preconditioning can be associated with any other preconditioner available in the package or defined by the end user since we provide generic interface similar to the other Krylov subspace methods in the package. The resulted KSP module (named as DGMRES) is available in PETSc release 3.2.

4 Numerical Experiments

This section presents some numerical results to prove the efficiency of the proposed approaches. The test problem arises from design optimization in computational fluid dynamics. The physical model is a 3D flow simulation in a jet engine compressor rotor. The physical equations are the Reynolds-Averaged Navier-Stokes for compressible flows, discretized using the finite volume method as presented by Aubert et al. [2]. The matrices have been extracted from the software Turb'Opty^{TM} designed by the FLUOREM company. They are also available in the University of

Florida sparse matrix collection (see [6]) under the name *RM07R* in the FLUOREM group. The matrix is nonsymmetric and indefinite with a size 272,635 and 37,355,908 nonzero entries. Other test cases can be found in [8].

With this test case so far, previous studies have shown the limits of some existing solvers in terms of memory usage and numerical accuracy (see [9]). Pacull et al. [10] have proved as well the instability of the ILU factorization to approximate the solution of linear subsystems. In our hybrid approach, we therefore rely on a direct solver within each subdomain, such as MUMPS [1].

4.1 Benefits of the Deflated Restarting

We now give the main benefits of using the deflated GMRES with the additive Schwarz method (ASM). It is known that one level ASM is a weak preconditioner when the number of subdomains D gets large. The size of the Krylov subspace m could then be increased to enhance the robustness of the global method. However, choosing a good size m of the Krylov subspace is a trial-and-error process. With the adaptive deflation, we show experimentally that the method is robust for various values of m and D. Moreover, using a large number of subdomains reduces the memory required to handle the submatrices by the direct solver. Hence it is expected that the time to factorize these matrices and the memory required will get smaller as D increases. This is reported in the last column of Table 1. We also report the number of matrix-vector multiplies and the global CPU time with respect to the number of subdomains D. We then compare the restarted version (GMRES(m)) with the deflated version (DGMRES(m,k)), where $m = 48$ and 64. A dash in a field means that the relative residual norm of 10^{-8} is not reached after 2500 iterations. It can be observed that DGMRES provides reliable and faster convergence than the classical restarted GMRES. It also gives a faster method since significantly fewer iterations are needed. Furthermore, the method reveals a substantial acceleration as the number of processors increases. Note that without the deflation, this acceleration will not be obtained since the number of matrix-vector multiplies increases hugely with the subdomains. For instance, this behavior can be seen with GMRES(64) when using $D = 16$ and $D = 32$.

Table 1. RM07R : Benefits of using DGMRES with an additive Schwarz preconditioner and an overlap of 1. The deflation process reduces the total number of iterations and helps to use a large number of subdomains and thus a large number of processors. Here, the number of processors is indeed equal to the number of subdomains.

D	GMRES(48)		DGMRES(47,1)			GMRES(64)		DGMRES(63,1)		
	Matvecs	Time	Matvecs	Time	r	Matvecs	Time	Matvecs	Time	r
16	551	230	212	173.4	3	355	193.8	208	168.9	2
32	-	-	533	109.2	4	2217	244.6	455	94.6	7
64	-	-	410	56.8	4	-	-	453	50.8	7
128	-	-	791	51.5	15	-	-	638	44.3	8

4.2 Adaptive DGMRES and Full GMRES

From the robustness standpoint, the full GMRES approach is more reliable than the restarted version even with the deflation process. However as the size of the basis grows, it should be more sensitive to round-off errors. To illustrate this behavior, we consider two formulations of the Arnoldi process, namely the classical Gram-Schmidt (*CGS*) and the modified Gram-Schmidt (*MGS*) algorithms. The former is sometimes preferred since it provides good kernel operations in parallel environments. In the PETSc package, for instance, it is used by default in the GMRES implementation as the orthogonalization method with a possible iterative refinement strategy. In Fig. 1, the residual history is displayed with respect to the number of matrix-vector products. The method stops when the relative residual norm is 10^{-10}. It can then be noticed that with *CGS*, stagnation occurs in the full GMRES (in solid line) due to severe cancellation in the algorithm and consequently a loss of orthogonality. This does not happen when the basis is small since the round-off errors are not propagated very far and DGMRES (dash-dotted line) converges at the desired accuracy even with *CGS*. Note that although good accuracy is finally achieved in

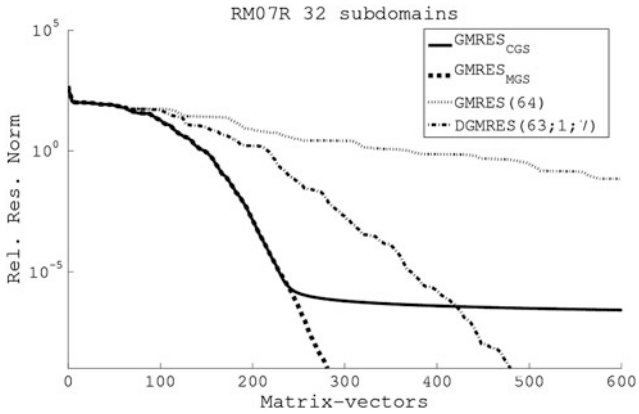


Fig. 1. Convergence of full GMRES, GMRES(m) and DGMRES(m,k,r) with classical Gram-Schmidt(*CGS*) and modified Gram-Schmidt (*MGS*) orthogonalization scheme. k is the number of eigenvalues to extract at each detected stagnation and r is the total number of eigenvalues extracted at the convergence. Thirty two subdomains are used in the additive Schwarz method with a 1-overlap

full GMRES with *MGS* (dashed line), it will require much more memory to store all the vectors of the growing Krylov basis (265 vectors in this case). In DGMRES, the Krylov basis is stored just for one cycle. Only the invariant basis U is stored over the restart cycles together with vectors $M^{-1}AU$ to reduce the matrix-vector counts. Thus in this example, only $63 + 7 \times 2 = 77$ vectors are stored. Note also that this number can be further reduced by using a smaller Krylov basis since convergence is still good, as shown in Table 1.

5 Conclusion

We have designed an adaptive deflation strategy that can be used for preconditioned GMRES. We show in this paper that the proposed algorithm can be used to improve the robustness and reduce both CPU time and memory required by hybrid solvers based on a one level additive Schwarz method. We have implemented this method in the new module DGMRES of the PETSc library.

Acknowledgments This work is funded by the French National Agency of Research under the contract ANR-TLOG07-011-03 LIBRAERO. The work of the first author was done while visiting the NCSA at Urbana-Champaign in the context of the Joint laboratory INRIA-University of Illinois. Experiments in this paper have been carried out using the *parapide* cluster in the GRID'5000 experimental testbed (see https://www.grid5000.fr). We thank the referees for providing many instructive comments.

Bibliography

[1] Patrick R. Amestoy, Iain S. Duff, Jean-Yves L'Excellent, and Jacko Koster. A fully asynchronous multifrontal solver using distributed dynamic scheduling. *SIAM J. Matrix Anal. Appl.*, 23(1):15–41 (electronic), 2001. ISSN 0895–4798.

[2] S. Aubert, J. Tournier, M. Rochette, J. Blanche, M. N'Diaye, S. Melen, M. Till, and P. Ferrand. Optimization of a gas mixer using a new parametric flow solver. In *Proceedings of the ECCOMAS Computational Fluid Dynamics Conference, Swansea, UK*, 2001.

[3] Satish Balay, Kris Buschelman, Victor Eijkhout, William D. Gropp, Dinesh Kaushik, Matthew G. Knepley, Lois Curfman McInnes, Barry F. Smith, and Hong Zhang. PETSc users manual. Technical Report ANL-95/11 - Revision 3.0.0, Argonne National Laboratory, 2008.

[4] Kevin Burrage and Jocelyne Erhel. On the performance of various adaptive preconditioned GMRES strategies. *Numer. Linear Algebra Appl.*, 5(2):101–121, 1998. ISSN 1070–5325.

[5] Xiao-Chuan Cai and Marcus Sarkis. A restricted additive Schwarz preconditioner for general sparse linear systems. *SIAM J. Sci. Comput.*, 21(2):792–797, 1999.

[6] Timothy A. Davis and Yifan Hu. The University of Florida Sparse Matrix Collection. *ACM Transactions on Mathematical Software*, 38(1), 2011.

[7] Jocelyne Erhel, Kevin Burrage, and Bert Pohl. Restarted GMRES preconditioned by deflation. *J. Comput. Appl. Math.*, 69(2):303–318, 1996. ISSN 0377–0427.

[8] Désiré Nuentsa Wakam and François Pacull. Memory efficient and robust hybrid algebraic solvers for large CFD linear systems. *Computer and Fluids*, submitted, 2011. special issue of ParCFD2011.

[9] Désiré Nuentsa Wakam, Jocelyne Erhel, Edouard Canot, and Guy-Antoine Atenekeng Kahou. A comparative study of some distributed linear solvers

on systems arising from fluid dynamics simulations. In *Parallel Computing: From Multicores and GPU's to Petascale*, volume 19 of *Advances in Parallel Computing*, pages 51–58. IOS Press, 2010.

[10] F. Pacull, S. Aubert, and M. Buisson. Study of ILU factorization for schwarz preconditioners with application to computational fluid dynamics. In *Proceedings of the Second International Conference on Parallel, Distributed, Grid and Cloud Computing for Engineering*. Civil-Comp Press, Stirlingshire, UK, 2011. doi: doi:10.4203/ccp.95.39.

[11] Youcef Saad and Martin H. Schultz. GMRES: a generalized minimal residual algorithm for solving nonsymmetric linear systems. *SIAM J. Sci. Stat. Comp.*, 7(3):856–869, 1986.

[12] Maria Sosonkina, Layne T. Watson, Rakesh K. Kapania, and Homer F. Walker. A new adaptive GMRES algorithm for achieving high accuracy. *Numer. Linear Algebra Appl.*, 5(4):275–297, 1998.

Quasi-optimality of BDDC Methods for MITC Reissner-Mindlin Problems

L. Beirão da Veiga[1], C. Chinosi[2], C. Lovadina[3], L.F. Pavarino[1], and J. Schöberl[4]

[1] Dipartimento di Matematica, Università di Milano, Via Saldini 50, 20133 Milano, Italy, lourenco.beirao@unimi.it, luca.pavarino@unimi.it
[2] Dipartimento di Scienze e Tecnologie Avanzate, Università del Piemonte Orientale, Via Bellini 25/G, I-15100 Alessandria, Italy, claudia.chinosi@mfn.unipmn.it
[3] Dipartimento di Matematica, Università di Pavia, Via Ferrata 1, 27100 Pavia, Italy, carlo.lovadina@unipv.it
[4] Institute for Analysis and Scientific Computing, Vienna University of Technology, Wiedner Hauptstrasse 8-10 1040 Wien, Austria, joachim.schoeberl@tuwien.ac.at

1 Introduction

The goal of this paper is to improve a condition number bound proven in [5] for a Balancing Domain Decomposition Method by Constraints (BDDC) for the Reissner-Mindlin plate bending problem discretized with MITC elements. This BDDC preconditioner is based on selecting the plate rotations and deflection degrees of freedom at the subdomain vertices as primal continuity constraints. In [5], we proved that the resulting BDDC algorithm is scalable in the number of subdomains N and independent of the plate thickness t and that the condition number κ of the preconditioned Reissner-Mindlin plate problem is bounded by

$$\kappa \leq C(H/h),$$

with C a constant independent of the plate thickness t, the mesh size h and the subdomain size H. In the present contribution, we prove the improved quasi-optimal result

$$\kappa \leq C(1+\log^3(H/h)).$$

We remark that the MITC discretization of Reissner-Mindlin problems can lead to very ill-conditioned discrete system, with condition number

$$\kappa_{no} \sim Ch^{-2}t^{-2}.$$

Introduced in [11] and analyzed in [17, 21, 22], BDDC methods have evolved from previous domain decomposition work on Balancing Neumann-Neumann methods. BDDC algorithm have been extended in recent years from scalar elliptic problems to almost incompressible elasticity [12, 24], the Stokes system [18], flow in porous

R. Bank et al. (eds.), *Domain Decomposition Methods in Science and Engineering XX*, Lecture Notes in Computational Science and Engineering 91,
DOI 10.1007/978-3-642-35275-1_76,

media [28], and spectral element discretizations [15, 23, 24]. BDDC and overlapping Schwarz methods for Reissner-Mindlin plate problems discretized with Falk-Tu elements have been studied in the recent Ph.D. thesis [16], while multigrid method for plates have been studied in [26]. Among the several finite element works for plates, we mention [2, 3, 7–10, 13, 14, 19, 20, 27].

2 The MITC Reissner-Mindlin Plate Bending Problem

Continuous problem. Let Ω be a polygonal domain in $\mathbb{R}^2$ representing the midsurface of the plate, for simplicity assumed to be clamped on the whole boundary $\partial\Omega$. The Reissner-Mindlin plate bending problem (see [1, 7]) reads

$$\begin{cases} \text{Find } \boldsymbol{\theta}^{ex} \in [H_0^1(\Omega)]^2, u^{ex} \in H_0^1(\Omega) \text{ such that} \\ a(\boldsymbol{\theta}^{ex}, \boldsymbol{\eta}) + \mu k t^{-2}(\boldsymbol{\theta}^{ex} - \nabla u^{ex}, \boldsymbol{\eta} - \nabla v) = (f, v) \quad \forall \boldsymbol{\eta} \in [H_0^1(\Omega)]^2, v \in H_0^1(\Omega)\,, \end{cases} \tag{1}$$

with μ the shear modulus, k is the shear correction factor, t the plate thickness, u^{ex} the deflection, $\boldsymbol{\theta}^{ex}$ the rotation of the normal fibers and f the applied scaled normal load. Moreover, $(\cdot,\cdot)$ stands for the standard scalar product in $L^2(\Omega)$ and $a(\cdot,\cdot)$ is the bilinear form

$$a(\boldsymbol{\theta}^{ex}, \boldsymbol{\eta}) = (\mathbb{C}\varepsilon(\boldsymbol{\theta}^{ex}), \varepsilon(\boldsymbol{\eta})),$$

with $\mathbb{C}$ the positive definite tensor of bending moduli and $\varepsilon(\cdot)$ the symmetric gradient operator. Introducing the scaled shear stresses $\boldsymbol{\gamma}^{ex} = \mu k t^{-2}(\boldsymbol{\theta}^{ex} - \nabla u^{ex})$, problem (1) can be written in terms of the following mixed variational formulation, where for simplicity we have assumed $\mu k = 1$:

$$\begin{cases} \text{Find } \boldsymbol{\theta}^{ex} \in [H_0^1(\Omega)]^2, u^{ex} \in H_0^1(\Omega), \boldsymbol{\gamma}^{ex} \in [L^2(\Omega)]^2 \text{ such that} \\ a(\boldsymbol{\theta}^{ex}, \boldsymbol{\eta}) + (\boldsymbol{\gamma}^{ex}, \boldsymbol{\eta} - \nabla v) = (f, v) \quad \forall \boldsymbol{\eta} \in [H_0^1(\Omega)]^2, v \in H_0^1(\Omega) \\ (\boldsymbol{\theta}^{ex} - \nabla u^{ex}, \boldsymbol{s}) - t^2(\boldsymbol{\gamma}, \boldsymbol{s}) = 0 \quad \forall \boldsymbol{s} \in [L^2(\Omega)]^2\,. \end{cases} \tag{2}$$

Discrete problem. We discretize the plate problem by MITC (Mixed Interpolation of Tensorial Components) elements; see e.g. [1, 7, 8] for more details on this family of elements. Let τ_h denote a triangular or quadrilateral conforming finite element mesh on Ω, of characteristic mesh size h. Let $\boldsymbol{\Theta}$, U and $\boldsymbol{\Gamma}$ be the discrete spaces for rotations, deflections and shear stresses, respectively and define $\mathbf{X} = \boldsymbol{\Theta} \times U$. Then the Reissner-Mindlin plate bending problem (2) discretized with MITC elements reads

$$\begin{cases} \text{Find } (\boldsymbol{\theta}, u) \in \mathbf{X},\ \boldsymbol{\gamma} \in \boldsymbol{\Gamma} \text{ such that} \\ a(\boldsymbol{\theta}, \boldsymbol{\eta}) + (\boldsymbol{\gamma}, \Pi\,\boldsymbol{\eta} - \nabla v) = (f, v) \quad \forall (\boldsymbol{\eta}, v) \in \mathbf{X} \\ (\Pi\,\boldsymbol{\theta} - \nabla u, \boldsymbol{s}) - t^2(\boldsymbol{\gamma}, \boldsymbol{s}) = 0 \quad \forall \boldsymbol{s} \in \boldsymbol{\Gamma}\,, \end{cases} \tag{3}$$

where $\Pi : ([H^1(\Omega)]^2 + \boldsymbol{\Gamma}) \longrightarrow \boldsymbol{\Gamma}$ is the MITC reduction operator. Using the second equation of (3), shear stresses can be eliminated to obtain the following positive definite discrete formulation:

$$\begin{cases} \text{Find } (\boldsymbol{\theta},u) \in \mathbf{X} \text{ such that} \\ b((\boldsymbol{\theta},u),(\boldsymbol{\eta},v)) = (f,v) \quad \forall (\boldsymbol{\eta},v) \in \mathbf{X}\,, \end{cases} \tag{4}$$

where we have defined $b((\boldsymbol{\theta},u),(\boldsymbol{\eta},v)) := a(\boldsymbol{\theta},\boldsymbol{\eta}) + t^{-2}(\Pi\,\boldsymbol{\theta} - \nabla u, \Pi\,\boldsymbol{\eta} - \nabla v)$. In this paper, we address directly the positive definite problem (4), in the spirit of [4, 5], instead of the mixed formulation (3). For the convergence analysis of the MITC elements, see e.g. [3, 8, 13, 25]. The MITC elements perform optimally with respect to the polynomial degree and regularity of the solution, and their rate of convergence is independent of the thickness parameter t.

3 Iterative Substructuring and BDDC Preconditioning

Subspace decomposition and Schur complement. We decompose the domain Ω into N open, nonoverlapping subdomains Ω_i of characteristic size H forming a shape-regular finite element mesh τ_H. This coarse triangulation τ_H is further refined into a finer triangulation τ_h of characteristic size h; both meshes will typically be composed of triangles or quadrilaterals. In the sequel, we assume that the material tensor $\mathbb{C}$ is constant on the whole domain.

As it is standard in iterative substructuring methods, we first reduce the problem to the interface $\Gamma = \left(\bigcup_{i=1}^{N} \partial\Omega_i\right) \setminus \partial\Omega$, by implicitly eliminating the interior degrees of freedom. In variational form, this process consists in a suitable decomposition of the discrete space $\mathbf{X} = \boldsymbol{\Theta} \times U$. More precisely, let us define $\boldsymbol{W} = \mathbf{X}_{|\Gamma}$, i.e. the space of the traces of functions in $\mathbf{X}$, as well as the local spaces $\mathbf{X}_i = \mathbf{X} \cap [H_0^1(\Omega_i)]^3$. The space $\mathbf{X}$ can be decomposed as $\mathbf{X} = \oplus_{i=1}^{N} \mathbf{X}_i \oplus \overline{\mathscr{H}}(\boldsymbol{W})$. Here $\overline{\mathscr{H}} : \boldsymbol{W} \longrightarrow \mathbf{X}$ is the discrete "plate-harmonic" extension operator defined by solving the problem

$$\begin{cases} \text{Find } \overline{\mathscr{H}}(\boldsymbol{w}_\Gamma) \in \mathbf{X} \text{ such that } \overline{\mathscr{H}}(\boldsymbol{w}_\Gamma)|_\Gamma = \boldsymbol{w}_\Gamma \text{ and} \\ b(\overline{\mathscr{H}}(\boldsymbol{w}_\Gamma), \boldsymbol{v}_I) = 0 \qquad \forall \boldsymbol{v}_I \in \mathbf{X}_i \quad i = 1,2,\ldots,N. \end{cases}$$

Defining the Schur complement bilinear form $s(\boldsymbol{w}_\Gamma, \boldsymbol{v}_\Gamma) = b(\overline{\mathscr{H}}(\boldsymbol{w}_\Gamma), \overline{\mathscr{H}}(\boldsymbol{v}_\Gamma))$, the Schur complement system reads $s(\boldsymbol{u}_\Gamma, \boldsymbol{v}_\Gamma) =< \tilde{\boldsymbol{f}}, \boldsymbol{v}_\Gamma > \ \forall \boldsymbol{v}_\Gamma \in \boldsymbol{W}$, for a suitable right-hand side $\tilde{\boldsymbol{f}}$.

The BDDC Reissner-Mindlin plate preconditioner. BDDC preconditioners, introduced in [11] and analyzed in [21], can be regarded as an evolution of Balancing Neumann-Neumann preconditioners for the Schur complement system. In this section, we briefly recall the BDDC preconditioner of [5].

Define $\Gamma_i := \partial\Omega_i$, and $\Gamma_{ij} = \partial\Omega_i \cap \partial\Omega_j$, $i,j \in \{1,2,\ldots,N\}$, the common edge between two adjacent subdomains Ω_i and Ω_j. The local spaces $\overline{\boldsymbol{W}}_i$ are the spaces of discrete functions defined by $\overline{\boldsymbol{W}}_i = \boldsymbol{W}_{|\Gamma_i}$, $i = 1,2,\ldots,N$. Let $\overline{\mathscr{H}}_i : \overline{\boldsymbol{W}}_i \longrightarrow \mathbf{X}|_{\Omega_i}$, $i = 1,2,\ldots,N$, represent the restriction of the operator $\overline{\mathscr{H}}$ to the subdomain Ω_i

$$\begin{cases} \text{Find } \overline{\mathscr{H}}_i(\boldsymbol{w}_i) \in \mathbf{X}|_{\Omega_i} \text{ such that } \overline{\mathscr{H}}(\boldsymbol{w}_i)|_{\Gamma_i} = \boldsymbol{w}_i \text{ and} \\ b_i(\overline{\mathscr{H}}_i(\boldsymbol{w}_i), \boldsymbol{v}_i) = 0 \qquad \forall \boldsymbol{v}_i \in \mathbf{X}_i, \end{cases}$$

where the $b_i(\cdot,\cdot)$ are given by restricting the integrals in $b(\cdot,\cdot)$ to the domain Ω_i, $i = 1,2,\ldots,N$. The local bilinear forms are $s_i(\boldsymbol{w}_i,\boldsymbol{v}_i) = b_i(\overline{\mathscr{H}}_i\boldsymbol{w}_i,\overline{\mathscr{H}}_i\boldsymbol{v}_i), \forall \boldsymbol{w}_i,\boldsymbol{v}_i \in \overline{\boldsymbol{W}}_i$. Let R_i^T, $i = 1,2,\ldots,N$ be the prolongation operators which extend any function of $\overline{\boldsymbol{W}}_i$ to the function of $\boldsymbol{W}$ which is zero at all the nodes not on Γ_i. Note that for $\boldsymbol{w},\boldsymbol{v} \in \boldsymbol{W}$, $\sum_{i=1}^N s_i(R_i\boldsymbol{w},R_i\boldsymbol{v}) = s(\boldsymbol{w},\boldsymbol{v})$. For $x \in \Gamma$, we also define the weight $N_x = \#\left\{j \in \mathbb{N} \,|\, x \in \partial\Omega_j\right\}$ and the weighted counting operators $\delta_i : \overline{\boldsymbol{W}}_i \longrightarrow \overline{\boldsymbol{W}}_i$ (and their inverses $\delta_i^\dagger$) by

$$\delta_i\boldsymbol{v}_i\,(x) = N_x\boldsymbol{v}_i(x), \qquad \delta_i^\dagger\boldsymbol{v}_i\,(x) = N_x^{-1}\boldsymbol{v}_i(x), \quad \forall x \text{ node of } \Gamma_i \cap \Gamma.$$

Let $C_i : \overline{\boldsymbol{W}}_i \to \mathbb{R}^{3cc_i}$ be local constraint operators that read function values at the corners of the subdomain Ω_i, with cc_i the number of corners of the subdomain. Then we define the local constrained spaces

$$\boldsymbol{W}_i = \{\boldsymbol{w}_i \in \overline{\boldsymbol{W}}_i \,|\, C_i\boldsymbol{w}_i = \boldsymbol{0}\},$$

and a global coarse space $\boldsymbol{W}_0 \subset \boldsymbol{W}$ associated with the function values at the subdomain vertices. Given the number m of such subdomain vertices, let $w_c \in \mathbb{R}^{3m}$ be a vector representing the respective nodal values. Then the space $\boldsymbol{W}_0$ is defined by

$$\boldsymbol{W}_0 = \{\sum_{i=1}^N R_i^T\,\delta_i^\dagger\boldsymbol{w}_{0,i} \,|\, C_i\boldsymbol{w}_{0,i} = R_i^C w_c, w_c \in \mathbb{R}^{3m}, s_i(\boldsymbol{w}_{0,i},\boldsymbol{w}_{0,i}) \to \min\},$$

with R_i^C the operator extracting the vertex values for the subdomain Ω_i from the global vector w_c of all the subdomain vertex values. Any element $\boldsymbol{w} \in \boldsymbol{W}$ can be uniquely decomposed as $\boldsymbol{w} = \boldsymbol{w}_0 + \sum_{i=1}^N \boldsymbol{w}_i$, with $\boldsymbol{w}_0 \in \boldsymbol{W}_0$, $\boldsymbol{w}_i \in \boldsymbol{W}_i$ for $i = 1,\ldots,N$. We use inexact bilinear forms defined by

$$\begin{aligned}
\tilde{s}_i(\boldsymbol{w}_i,\boldsymbol{v}_i) &= s_i(\delta_i\boldsymbol{w}_i,\delta_i\boldsymbol{v}_i) \qquad \forall \boldsymbol{w}_i,\boldsymbol{v}_i \in \boldsymbol{W}_i,\ i = 1,2,\ldots,N,\\
\tilde{s}_0(\boldsymbol{w}_0,\boldsymbol{v}_0) &= \sum_{i=1}^N s_i(\boldsymbol{w}_{0,i},\boldsymbol{v}_{0,i}) \qquad \forall \boldsymbol{w}_0,\boldsymbol{v}_0 \in \boldsymbol{W}_0.
\end{aligned}$$

Finally, we define the coarse operator $P_0 : \boldsymbol{W} \longrightarrow \boldsymbol{W}_0$ by

$$\tilde{s}_0(P_0\boldsymbol{u},\boldsymbol{v}_0) = s(\boldsymbol{u},\boldsymbol{v}_0) \ \ \forall \boldsymbol{v}_0 \in \boldsymbol{W}_0,$$

and the local operators $P_i = R_i^T\tilde{P}_i : \boldsymbol{W} \longrightarrow R_i^T\boldsymbol{W}_i$ by

$$\tilde{s}_i(\tilde{P}_i\boldsymbol{u},\boldsymbol{v}_i) = s(\boldsymbol{u},R_i^T\boldsymbol{v}_i) \ \ \forall \boldsymbol{v}_i \in \boldsymbol{W}_i.$$

Then, our BDDC method is defined by the preconditioned operator

$$P = \sum_{i=0}^N P_i\,. \tag{5}$$

The matrix form of P and the associated preconditioner can be found in [5].

4 A Quasi-optimal BDDC Convergence Bound

We start by recalling the following assumption from [5], using the same notations.

Assumption 4 *Given any Γ_i, $i = 1,2,\ldots,N$, let $\mathscr{E}_i$ represent the set of the edges of Γ_i. Then, we assume that there exist two positive constants k_*, k^* and a boundary seminorm $|\cdot|_{\tau(\Gamma_i)}$ on $\overline{\boldsymbol{W}}_i$, $i = 1,2,\ldots,N$, such that*

$$|\boldsymbol{w}_i|^2_{\tau(\Gamma_i)} \leq k^* s_i(\boldsymbol{w}_i,\boldsymbol{w}_i) \quad \forall \boldsymbol{w}_i \in \overline{\boldsymbol{W}}_i, \tag{6}$$

$$|\boldsymbol{w}_i|^2_{\tau(\Gamma_i)} \geq k_* s_i(\boldsymbol{w}_i,\boldsymbol{w}_i) \quad \forall \boldsymbol{w}_i \in \boldsymbol{W}_i, \tag{7}$$

$$|\boldsymbol{w}_i|^2_{\tau(\Gamma_i)} = \sum_{e\in\mathscr{E}_i} |\boldsymbol{w}_i|^2_{\tau(e)} \quad \forall \boldsymbol{w}_i \in \overline{\boldsymbol{W}}_i, \tag{8}$$

where $|\cdot|_{\tau(e)}$ is a given seminorm on the edge e.

We notice that we cannot adopt the obvious choice $|\boldsymbol{w}_i|_{\tau(\Gamma_i)} = s_i(\boldsymbol{w}_i,\boldsymbol{w}_i)$, since it can be shown that it does not satisfy (8), not even with a bound including a uniform constant. We have the following main result.

Theorem 2. *If Assumption 4 holds, then the condition number κ of the Reissner-Mindlin BDDC preconditioned operator P in (5) satisfies the bound*

$$\kappa(P) \leq C\left(1+\log^3(H/h)\right),$$

with the constant C depending only on the material constants and mesh regularity, and not on the plate thickness t.

Here we can only outline the main steps of the proof; full details can be found in [6]. The proof proceeds by showing that Assumption 4 holds for the MITC plate bending problem (4) and by establishing the respective upper and lower bounds for the constants k_*, k^* in (6), (7). These bounds in turn will prove Theorem 2 since $\kappa(P) \leq C(1+5k_*^{-1}k^*)$, see [5, 21] for a proof.

Upper bound (6). The upper bound is established exactly as in [5, Sect. 5.2].

Lower bound (7). To prove the lower bound, we note that the local spaces $\overline{\boldsymbol{W}}_i$, $i = 1,2,\ldots,N$, are composed of rotation and deflection parts, which we denote by $\overline{\boldsymbol{W}}_i = \overline{\boldsymbol{\Theta}}_i \times \overline{U}_i$. Accordingly, we denote the rotation and deflection parts of the constrained space by $\boldsymbol{W}_i = \boldsymbol{\Theta}_i \times U_i$, where the functions of $\boldsymbol{\Theta}_i$ and U_i vanish at the subdomain corner nodes. We work with the following seminorm defined in [5]: $|\boldsymbol{w}_i|^2_{\tau(\Gamma_i)} = \sum_{e\in\mathscr{E}_i} |\boldsymbol{w}_i|^2_{\tau(e)} \quad \forall \boldsymbol{w}_i = (\boldsymbol{\theta}_i, u_i) \in \overline{\boldsymbol{W}}_i$, where for all edges $e \in \mathscr{E}_i$

$$|\boldsymbol{w}_i|^2_{\tau(e)} = |\boldsymbol{\theta}_i|^2_{\gamma(e)} + ht^{-2}||\Pi\,\boldsymbol{\theta}_i\cdot\boldsymbol{\tau} - u_i'||^2_{L^2(e)},$$

$$|\boldsymbol{\theta}_i|_{\gamma(e)} := \inf_{\boldsymbol{\psi}\in[H^1(\Omega_i)]^2, \boldsymbol{\psi}|_e = \boldsymbol{\theta}_i|_e} ||\boldsymbol{\varepsilon}(\boldsymbol{\psi})||_{L^2(\Omega_i)},$$

$\boldsymbol{\tau}$ is the tangent unit vector at the boundary and the apex indicates the derivative, in the direction of $\boldsymbol{\tau}$, for functions defined on the (one dimensional) boundary. We

now improve the lower bound proved in [5] by introducing a splitting of the plate rotation variable. Consider $\boldsymbol{w}_i = (\boldsymbol{\theta}_i, u_i) \in \boldsymbol{W}_i$ and define the splitting $\boldsymbol{\theta}_i^{(2)} \in \boldsymbol{\Theta}_i^{(2)} := \text{span}\{B_l^i \boldsymbol{\tau}\}_{l \in \Gamma_i}$, by

$$\int_e \boldsymbol{\theta}_i^{(2)} \cdot \boldsymbol{\tau} = \int_e \boldsymbol{\theta}_i \cdot \boldsymbol{\tau} - u_i' \quad \forall e \in \mathscr{E}_i,$$

and let $\boldsymbol{\theta}_i^{(1)} = \boldsymbol{\theta}_i - \boldsymbol{\theta}_i^{(2)}$ so that $\boldsymbol{\theta}_i = \boldsymbol{\theta}_i^{(1)} + \boldsymbol{\theta}_i^{(2)}$. By construction, it holds

$$\int_e u_i' - \boldsymbol{\theta}_i^{(1)} \cdot \boldsymbol{\tau} = 0 \quad \forall e \in \mathscr{E}_i.$$

We introduce also the related splitting of $\boldsymbol{w}_i$

$$\boldsymbol{w}_i = \boldsymbol{w}_i^{(1)} + \boldsymbol{w}_i^{(2)}, \qquad \boldsymbol{w}_i^{(1)} = (u_i, \boldsymbol{\theta}_i^{(1)}), \qquad \boldsymbol{w}_i^{(2)} = (0, \boldsymbol{\theta}_i^{(2)}).$$

An improved lower bound can be obtained by estimating the split terms in the following two lemmas; see [6] for complete proofs.

Lemma 1. *There exists a constant $C > 0$ independent of h such that for all edges e of all subdomains Ω_i*

$$|\boldsymbol{w}_i|_{\tau(e)} = |(u_i, \boldsymbol{\theta}_i)|_{\tau(e)} \geq C\big(|(u_i, \boldsymbol{\theta}_i^{(1)})|_{\tau(e)} + |(0, \boldsymbol{\theta}_i^{(2)})|_{\tau(e)}\big).$$

This lemma follows from the inequality $||(0, \boldsymbol{\theta}_i^{(2)})||_{\tau(e)} \leq C||\boldsymbol{w}_i||_{\tau(e)}$, that is derived in [6] from the definition of $\boldsymbol{\theta}_i^{(2)}$, a scaling argument and an inverse inequality. A similar argument applied to the extension of $\boldsymbol{\theta}_i^2$ by zero inside Ω_i leads to the following lemma.

Lemma 2. *There exists a constant $C > 0$ independent of h such that*

$$s_i(\boldsymbol{w}_i^{(2)}, \boldsymbol{w}_i^{(2)}) \leq C\, |\boldsymbol{w}_i^{(2)}|^2_{\tau(\Gamma_i)}.$$

The main step in the proof of Theorem 2 is the bound of the following proposition, obtained by considering an auxiliary rotated Stokes problem with boundary data $\boldsymbol{\theta}_i^{(1)}$ and several technical estimates, see [6, Proposition 5.5].

Proposition 1. *There exists a constant $C > 0$ independent of h such that*

$$s_i(\boldsymbol{w}_i^{(1)}, \boldsymbol{w}_i^{(1)}) \leq C\,(1 + \log^3(H/h))\, |\boldsymbol{w}_i^{(1)}|^2_{\tau(\Gamma_i)}.$$

The upper bound then follows by combining the three previous results. Indeed, first recalling the splitting $\boldsymbol{w}_i = \boldsymbol{w}_i^{(1)} + \boldsymbol{w}_i^{(2)}$ and using a triangle inequality, then applying Lemma 2 and Proposition 1, finally using Lemma 1 yields

$$s_i(\boldsymbol{w}_i, \boldsymbol{w}_i) \leq 2\Big(s_i(\boldsymbol{w}_i^{(1)}, \boldsymbol{w}_i^{(1)}) + s_i(\boldsymbol{w}_i^{(2)}, \boldsymbol{w}_i^{(2)})\Big)$$

$$\leq C\Big((1 + \log^3(H/h))\, |\boldsymbol{w}_i^{(1)}|^2_{\tau(\Gamma_i)} + |\boldsymbol{w}_i^{(2)}|^2_{\tau(\Gamma_i)}\Big) \leq C(1 + \log^3(H/h))|\boldsymbol{w}_i|^2_{\tau(\Gamma_i)}.$$

Bound (7) is therefore proved with $k_*^{-1} = C\,(1 + \log^3(H/h))$, with the constant C depending only on the material constants and mesh regularity.

We remark that an extensive set of numerical tests, also including jump in the coefficients, which are in complete accordance with Theorem 2, can be found in [5].

Bibliography

[1] K. J. Bathe. *Finite Element Procedures in Engineering Analysis*. Prentice-Hall, Englewood Cliffs, NJ, 1982.

[2] L. Beirão da Veiga. Finite element methods for a modified Reissner-Mindlin free plate model. *SIAM J. Numer. Anal.*, 42(4):1572–1591, 2004.

[3] L. Beirão da Veiga. Optimal error bounds for the MITC4 plate bending element. *Calcolo*, 41(4):227–245, 2004.

[4] L. Beirão da Veiga, C. Lovadina, and L. F. Pavarino. Positive definite balancing Neumann-Neumann preconditioners for nearly incompressible elasticity. *Numer. Math.*, 104(3):271–296, 2006.

[5] L. Beirão da Veiga, C. Chinosi, C. Lovadina, and L. F. Pavarino. Robust BDDC preconditioners for Reissner-Mindlin plate bending problems and MITC elements. *SIAM J. Numer. Anal.*, 47(6):4214–4238, 2010.

[6] L. Beirão da Veiga, C. Chinosi, C. Lovadina, L. F. Pavarino, and J. Schöberl. Quasi-uniformity of BDDC preconditioners for the MITC Reissner-Mindlin problem. Technical Report 4PV11/2/0, I.M.A.T.I.-C.N.R., 2011.

[7] F. Brezzi and M. Fortin. *Mixed and hybrid finite element methods*. Springer-Verlag, New York, 1991.

[8] F. Brezzi, M. Fortin, and R. Stenberg. Error analysis of mixed-interpolated elements for Reissner-Mindlin plates. *Math. Models Methods Appl. Sci.*, 1(2): 125–151, 1991.

[9] D. Chapelle and R. Stenberg. An optimal low-order locking-free finite element method for Reissner-Mindlin plates. *Math. Models Methods Appl. Sci.*, 8(3): 407–430, 1998.

[10] C. Chinosi, C. Lovadina, and L. D. Marini. Nonconforming locking-free finite elements for Reissner-Mindlin plates. *Comput. Methods Appl. Mech. Engrg.*, 195(25–28):3448–3460, 2006.

[11] C. R. Dohrmann. A preconditioner for substructuring based on constrained energy minimization. *SIAM J. Sci. Comput.*, 25(1):246–258, 2003.

[12] C. R. Dohrmann. A substructuring preconditioner for nearly incompressible elasticity problems. Technical Report SAND2004–5393, Sandia National Laboratories, 2004.

[13] R. Durán and E. Liberman. On mixed finite element methods for the Reissner-Mindlin plate model. *Math. Comp.*, 58(198):561–573, 1992.

[14] R. S. Falk and T. Tu. Locking-free finite elements for the Reissner-Mindlin plate. *Math. Comp.*, 69(231):911–928, 2000.

[15] A. Klawonn, L. F. Pavarino, and O. Rheinbach. Spectral element FETI-DP and BDDC preconditioners with multi-element subdomains. *Comput. Methods Appl. Mech. Engrg.*, 198(3–4):511–523, 2008.

[16] J. H. Lee. *Domain Decomposition Methods for Reissner-Mindlin Plates discretized with the Falk-Tu Elements*. PhD thesis, Courant Institute, NYU, 2011.

[17] J. Li and O. B. Widlund. FETI-DP, BDDC, and block Cholesky methods. *Internat. J. Numer. Methods Engrg.*, 66(2):250–271, 2006.

[18] J. Li and O.B. Widlund. BDDC algorithms for incompressible Stokes equations. *SIAM J. Numer. Anal.*, 44(6):2432–2455, 2006.

[19] C. Lovadina. A new class of mixed finite element methods for Reissner-Mindlin plates. *SIAM J. Numer. Anal.*, 33(6):2457–2467, 1996.

[20] C. Lovadina. A low-order nonconforming finite element for Reissner-Mindlin plates. *SIAM J. Numer. Anal.*, 42(6):2688–2705, 2005.

[21] J. Mandel and C. R. Dohrmann. Convergence of a balancing domain decomposition by constraints and energy minimization. *Numer. Linear Algebra Appl.*, 10(7):639–659, 2003. Dedicated to the 70th birthday of Ivo Marek.

[22] J. Mandel, C. R. Dohrmann, and R. Tezaur. An algebraic theory for primal and dual substructuring methods by constraints. *Appl. Numer. Math.*, 54(2): 167–193, 2005.

[23] L. F. Pavarino. BDDC and FETI-DP preconditioners for spectral element discretizations. *Comput. Meth. Appl. Mech. Engrg.*, 196(8):1380–1388, 2007.

[24] L. F. Pavarino, Widlund O. B., and Zampini S. BDDC preconditioners for spectral element discretizations of almost incompressible elasticity in three dimensions. *SIAM J. Sci. Comput.*, 32(6):3604–3626, 2010.

[25] J. Pitkäranta and M. Suri. Upper and lower error bounds for plate-bending finite elements. *Numer. Math.*, 84(4):611–648, 2000.

[26] J. Schöberl and R. Stenberg. Multigrid methods for a stabilized Reissner-Mindlin plate formulation. *SIAM J. Numer. Anal.*, 47(4):2735–2751, 2009.

[27] R. Stenberg. A new finite element formulation for the plate bending problem. In *Asymptotic methods for elastic structures (Lisbon, 1993)*, pages 209–221. de Gruyter, Berlin, 1995.

[28] X. Tu. A BDDC algorithm for a mixed formulation of flow in porous media. *Electron. Trans. Numer. Anal.*, 20:164–179, 2005.

Penalty Robin-Robin Domain Decomposition Schemes for Contact Problems of Nonlinear Elasticity

Ihor I. Prokopyshyn[1], Ivan I. Dyyak[2], Rostyslav M. Martynyak[1], and Ivan A. Prokopyshyn[2]

[1] Pidstryhach IAPMM NASU, Naukova 3-b, Lviv, 79060, Ukraine, ihor84@gmail.com
[2] Ivan Franko National University of Lviv, Universytetska 1, Lviv, 79000, Ukraine

1 Introduction

Many domain decomposition techniques for contact problems have been proposed on discrete level, particularly substructuring and FETI methods [1, 4].

Domain decomposition methods (DDMs), presented in [2, 10, 11, 16] for unilateral two-body contact problems of linear elasticity, are obtained on continuous level. All of them require to solve the nonlinear one-sided contact problems for one or both of the bodies in each iteration.

In works [6, 14, 15] we have proposed a class of penalty parallel Robin–Robin domain decomposition schemes for unilateral multibody contact problems of linear elasticity, which are based on penalty method and iterative methods for nonlinear variational equations. In each iteration of these schemes we have to solve in a parallel way some linear variational equations in subdomains.

In this contribution we generalize domain decomposition schemes, proposed in [6, 14, 15] to the solution of unilateral and ideal multibody contact problems of nonlinear elasticity. We also present theorems about the convergence of these schemes.

2 Formulation of Multibody Contact Problem

Consider a contact of N nonlinear elastic bodies $\Omega_\alpha \subset \mathbb{R}^3$ with sectionally smooth boundaries Γ_α, $\alpha = 1,2,\ldots,N$ (Fig. 1). Denote $\Omega = \bigcup_{\alpha=1}^{N} \Omega_\alpha$.

A stress-strain state in point $\mathbf{x} = (x_1, x_2, x_3)^\top$ of each body Ω_α is defined by the displacement vector $\mathbf{u}_\alpha = u_{\alpha i}\,\mathbf{e}_i$, the tensor of strains $\hat{\boldsymbol{\varepsilon}}_\alpha = \varepsilon_{\alpha ij}\,\mathbf{e}_i\mathbf{e}_j$ and the tensor of stresses $\hat{\boldsymbol{\sigma}}_\alpha = \sigma_{\alpha ij}\,\mathbf{e}_i\mathbf{e}_j$. These quantities satisfy Cauchy relations, equilibrium equations and nonlinear stress-strain law [8]:

$$\sigma_{\alpha ij} = \lambda_\alpha\,\delta_{ij}\,\Theta_\alpha + 2\,\mu_\alpha\,\varepsilon_{\alpha ij} - 2\,\mu_\alpha\,\omega_\alpha(e_\alpha)\,e_{\alpha ij}\,,\ \ i,j = 1,2,3\,, \tag{1}$$

where $\Theta_\alpha = \varepsilon_{\alpha 11} + \varepsilon_{\alpha 22} + \varepsilon_{\alpha 33}$ is the volume strain, $\lambda_\alpha(\mathbf{x}) > 0$, $\mu_\alpha(\mathbf{x}) > 0$ are bounded Lame parameters, $e_{\alpha ij} = \varepsilon_{\alpha ij} - \delta_{ij}\,\Theta_\alpha/3$ are the components of the strain

R. Bank et al. (eds.), *Domain Decomposition Methods in Science and Engineering XX*, Lecture Notes in Computational Science and Engineering 91,
DOI 10.1007/978-3-642-35275-1_77,

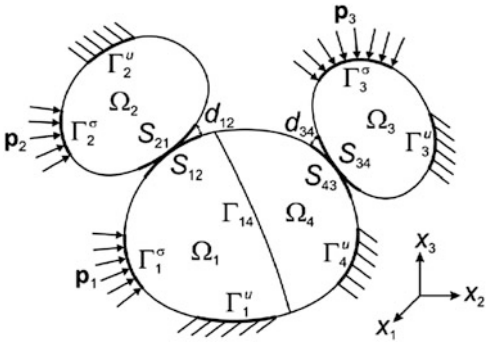


Fig. 1. Contact of several bodies

deviation tensor, $e_\alpha = \sqrt{2g_\alpha}/3$ is the deformation intensity, $g_\alpha = (\varepsilon_{\alpha 11} - \varepsilon_{\alpha 22})^2 + (\varepsilon_{\alpha 22} - \varepsilon_{\alpha 33})^2 + (\varepsilon_{\alpha 33} - \varepsilon_{\alpha 11})^2 + 6(\varepsilon_{\alpha 12}^2 + \varepsilon_{\alpha 23}^2 + \varepsilon_{\alpha 31}^2)$, and $\omega_\alpha(z)$ is nonlinear differentiable function, which satisfies the following properties:

$$0 \le \omega_\alpha(z) \le \partial(z\,\omega_\alpha(z))/\partial z < 1\,,\ \ \partial(\omega_\alpha(z))/\partial z \ge 0\,. \tag{2}$$

On the boundary Γ_α let us introduce the local orthonormal basis $\boldsymbol{\xi}_\alpha$, $\boldsymbol{\eta}_\alpha$, $\mathbf{n}_\alpha$, where $\mathbf{n}_\alpha$ is the outer unit normal to Γ_α. Then the vectors of displacements and stresses on the boundary can be written in the following way: $\mathbf{u}_\alpha = u_{\alpha\xi}\boldsymbol{\xi}_\alpha + u_{\alpha\eta}\boldsymbol{\eta}_\alpha + u_{\alpha n}\mathbf{n}_\alpha$, $\boldsymbol{\sigma}_\alpha = \hat{\boldsymbol{\sigma}}_\alpha \cdot \mathbf{n}_\alpha = \sigma_{\alpha\xi}\boldsymbol{\xi}_\alpha + \sigma_{\alpha\eta}\boldsymbol{\eta}_\alpha + \sigma_{\alpha n}\mathbf{n}_\alpha$.

Suppose that the boundary Γ_α of each body consists of four disjoint parts: $\Gamma_\alpha = \Gamma_\alpha^u \bigcup \Gamma_\alpha^\sigma \bigcup \Gamma_\alpha^I \bigcup S_\alpha$, $\Gamma_\alpha^u \neq \emptyset$, $\Gamma_\alpha^u = \overline{\Gamma_\alpha^u}$, $\Gamma_\alpha^I \bigcup S_\alpha \neq \emptyset$, where $S_\alpha = \bigcup_{\beta \in B_\alpha} S_{\alpha\beta}$, and $\Gamma_\alpha^I = \bigcup_{\beta' \in I_\alpha} \Gamma_{\alpha\beta'}$. Surface $S_{\alpha\beta}$ is the possible unilateral contact area of body Ω_α with body Ω_β, and $B_\alpha \subset \{1,2,\ldots,N\}$ is the set of the indices of all bodies in unilateral contact with body Ω_α. Surface $\Gamma_{\alpha\beta'} = \Gamma_{\beta'\alpha}$ is the ideal contact area between bodies Ω_α and $\Omega_{\beta'}$, and $I_\alpha \subset \{1,2,\ldots,N\}$ is the set of the indices of all bodies which have ideal contact with Ω_α.

We assume that the areas $S_{\alpha\beta} \subset \Gamma_\alpha$ and $S_{\beta\alpha} \subset \Gamma_\beta$ are sufficiently close ($S_{\alpha\beta} \approx S_{\beta\alpha}$), and $\mathbf{n}_\alpha(\mathbf{x}) \approx -\mathbf{n}_\beta(\mathbf{x}')$, $\mathbf{x} \in S_{\alpha\beta}$, $\mathbf{x}' = P(\mathbf{x}) \in S_{\beta\alpha}$, where $P(\mathbf{x})$ is the projection of $\mathbf{x}$ on $S_{\alpha\beta}$ [12]. Let $d_{\alpha\beta}(\mathbf{x}) = \pm\|\mathbf{x} - \mathbf{x}'\|_2$ be a distance between bodies Ω_α and Ω_β before the deformation. The sign of $d_{\alpha\beta}$ depends on a statement of the problem.

We consider homogenous Dirichlet boundary conditions on the part Γ_α^u, and Neumann boundary conditions on the part Γ_α^σ:

$$\mathbf{u}_\alpha(\mathbf{x}) = 0,\ \mathbf{x} \in \Gamma_\alpha^u;\ \ \boldsymbol{\sigma}_\alpha(\mathbf{x}) = \mathbf{p}_\alpha(\mathbf{x}),\ \mathbf{x} \in \Gamma_\alpha^\sigma\,. \tag{3}$$

On the possible contact areas $S_{\alpha\beta}$, $\beta \in B_\alpha$, $\alpha = 1,2,\ldots,N$ the following nonlinear unilateral contact conditions hold:

$$\sigma_{\alpha n}(\mathbf{x}) = \sigma_{\beta n}(\mathbf{x}') \le 0\,,\ \ \sigma_{\alpha\xi}(\mathbf{x}) = \sigma_{\beta\xi}(\mathbf{x}') = \sigma_{\alpha\eta}(\mathbf{x}) = \sigma_{\beta\eta}(\mathbf{x}') = 0\,, \tag{4}$$

$$u_{\alpha n}(\mathbf{x}) + u_{\beta n}(\mathbf{x}') \le d_{\alpha\beta}(\mathbf{x})\,, \tag{5}$$

$$\big(u_{\alpha n}(\mathbf{x}) + u_{\beta n}(\mathbf{x}') - d_{\alpha\beta}(\mathbf{x})\big)\,\sigma_{\alpha n}(\mathbf{x}) = 0\,,\ \mathbf{x} \in S_{\alpha\beta}\,,\ \mathbf{x}' = P(\mathbf{x}) \in S_{\beta\alpha}\,. \tag{6}$$

On ideal contact areas $\Gamma_{\alpha\beta'} = \Gamma_{\beta'\alpha}$, $\beta' \in I_\alpha$, $\alpha = 1,2,\ldots,N$ we consider ideal mechanical contact conditions:

$$\mathbf{u}_\alpha(\mathbf{x}) = \mathbf{u}_{\beta'}(\mathbf{x})\,,\ \boldsymbol{\sigma}_\alpha(\mathbf{x}) = -\boldsymbol{\sigma}_{\beta'}(\mathbf{x}),\ \mathbf{x} \in \Gamma_{\alpha\beta'}\,. \tag{7}$$

3 Penalty Variational Formulation of the Problem

For each body Ω_α consider Sobolev space $V_\alpha = [H^1(\Omega_\alpha)]^3$ and the closed subspace $V_\alpha^0 = \{\mathbf{u}_\alpha \in V_\alpha :\ \mathbf{u}_\alpha = 0 \text{ on } \Gamma_\alpha^u\}$. All values of the elements from spaces V_α and V_α^0 on the parts of boundary Γ_α should be understood as traces [9].

Define Hilbert space $V_0 = V_1^0 \times \ldots \times V_N^0$ with the scalar product $(\mathbf{u},\mathbf{v})_{V_0} = \sum_{\alpha=1}^N (\mathbf{u}_\alpha,\mathbf{v}_\alpha)_{V_\alpha}$ and norm $\|\mathbf{u}\|_{V_0} = \sqrt{(\mathbf{u},\mathbf{u})_{V_0}}$, $\mathbf{u},\mathbf{v} \in V_0$. Introduce the closed convex set of all displacements in V_0, which satisfy nonpenentration contact conditions (5) and ideal kinematic contact conditions:

$$K = \big\{\mathbf{u} \in V_0 :\ u_{\alpha n} + u_{\beta n} \le d_{\alpha\beta} \text{ on } S_{\alpha\beta},\ \mathbf{u}_{\alpha'} = \mathbf{u}_{\beta'} \text{ on } \Gamma_{\alpha'\beta'}\big\}\,, \tag{8}$$

where $\{\alpha,\beta\} \in Q$, $Q = \{\{\alpha,\beta\} :\ \alpha \in \{1,2,\ldots,N\}\,,\ \beta \in B_\alpha\}$, $\{\alpha',\beta'\} \in Q^I$, $Q^I = \{\{\alpha',\beta'\} :\ \alpha' \in \{1,2,\ldots,N\}\,,\ \beta' \in I_\alpha\}$, and $d_{\alpha\beta} \in H_{00}^{1/2}(\Xi_\alpha)$, $\Xi_\alpha = \operatorname{int}(\Gamma_\alpha \setminus \Gamma_\alpha^u)$.

Let us introduce bilinear form $A(\mathbf{u},\mathbf{v}) = \sum_{\alpha=1}^N a_\alpha(\mathbf{u}_\alpha,\mathbf{v}_\alpha)$, $\mathbf{u},\mathbf{v} \in V_0$, which represents the total elastic deformation energy of the system of bodies, linear form $L(\mathbf{v}) = \sum_{\alpha=1}^N l_\alpha(\mathbf{v}_\alpha)$, $\mathbf{v} \in V_0$, which is equal to the external forces work, and nonquadratic functional $H(\mathbf{v}) = \sum_{\alpha=1}^N h_\alpha(\mathbf{v}_\alpha)$, $\mathbf{v} \in V_0$, which represents the total nonlinear deformation energy:

$$a_\alpha(\mathbf{u}_\alpha,\mathbf{v}_\alpha) = \int_{\Omega_\alpha} \big[\,\lambda_\alpha \Theta_\alpha(\mathbf{u}_\alpha)\,\Theta_\alpha(\mathbf{v}_\alpha) + 2\mu_\alpha \sum_{i,j} \varepsilon_{\alpha ij}(\mathbf{u}_\alpha)\,\varepsilon_{\alpha ij}(\mathbf{v}_\alpha)\,\big]\,d\Omega\,, \tag{9}$$

$$l_\alpha(\mathbf{v}_\alpha) = \int_{\Omega_\alpha} \mathbf{f}_\alpha \cdot \mathbf{v}_\alpha\,d\Omega + \int_{\Gamma_\alpha^\sigma} \mathbf{p}_\alpha \cdot \mathbf{v}_\alpha\,dS\,, \tag{10}$$

$$h_\alpha(\mathbf{v}_\alpha) = 3\int_{\Omega_\alpha} \mu_\alpha \int_0^{e_\alpha(\mathbf{v}_\alpha)} z\,\omega_\alpha(z)\,dz\,d\Omega\,, \tag{11}$$

where $\mathbf{p}_\alpha \in [H_{00}^{-1/2}(\Xi_\alpha)]^3$, and $\mathbf{f}_\alpha \in [L_2(\Omega_\alpha)]^3$ is the vector of volume forces.

Using [12], we have shown that the original contact problem has an alternative weak formulation as the following minimization problem on the set K:

$$F(\mathbf{u}) = A(\mathbf{u},\mathbf{u})/2 - H(\mathbf{u}) - L(\mathbf{u}) \to \min_{\mathbf{u} \in K}. \tag{12}$$

Bilinear form A is symmetric, continuous with constant $M_A > 0$ and coercive with constant $B_A > 0$, and linear form L is continuous. Nonquadratic functional H is doubly Gateaux differentiable in V_0:

$$H'(\mathbf{u},\mathbf{v}) = \sum_{\alpha} h'_{\alpha}(\mathbf{u}_{\alpha},\mathbf{v}_{\alpha}),\ H''(\mathbf{u},\mathbf{v},\mathbf{w}) = \sum_{\alpha} h''_{\alpha}(\mathbf{u}_{\alpha},\mathbf{v}_{\alpha},\mathbf{w}_{\alpha}),\ \mathbf{u},\mathbf{v},\mathbf{w} \in V_0, \tag{13}$$

$$h'_{\alpha}(\mathbf{u}_{\alpha},\mathbf{v}_{\alpha}) = 2\int_{\Omega_{\alpha}} \mu_{\alpha}\,\omega_{\alpha}(e_{\alpha}(\mathbf{u}_{\alpha}))\sum_{i,j} e_{\alpha ij}(\mathbf{u}_{\alpha})\,e_{\alpha ij}(\mathbf{v}_{\alpha})\,d\Omega. \tag{14}$$

Moreover, we have proved that the following conditions hold:

$$(\exists C>0)\,(\forall \mathbf{u}\in V_0)\,\{(1-C)A(\mathbf{u},\mathbf{u}) \geq 2H(\mathbf{u})\}, \tag{15}$$

$$(\forall \mathbf{u}\in V_0)\,(\exists R>0)\,(\forall \mathbf{v}\in V_0)\left\{\left|H'(\mathbf{u},\mathbf{v})\right| \leq R\,\|\mathbf{v}\|_{V_0}\right\}, \tag{16}$$

$$(\exists D>0)\,(\forall \mathbf{u},\mathbf{v},\mathbf{w}\in V_0)\left\{\left|H''(\mathbf{u},\mathbf{v},\mathbf{w})\right| \leq D\,\|\mathbf{v}\|_{V_0}\,\|\mathbf{w}\|_{V_0}\right\}, \tag{17}$$

$$(\exists B>0)\,(\forall \mathbf{u},\mathbf{v}\in V_0)\left\{A(\mathbf{v},\mathbf{v}) - H''(\mathbf{u},\mathbf{v},\mathbf{v}) \geq B\,\|\mathbf{v}\|^2_{V_0}\right\}. \tag{18}$$

From these properties, it follows that there exists a unique solution $\bar{\mathbf{u}} \in K$ of minimization problem (12), and this problem is equivalent to the following variational inequality on the set K:

$$A(\mathbf{u},\mathbf{v}-\mathbf{u}) - H'(\mathbf{u},\mathbf{v}-\mathbf{u}) - L(\mathbf{v}-\mathbf{u}) \geq 0,\ \forall \mathbf{v}\in K,\ \mathbf{u}\in K. \tag{19}$$

To obtain a minimization problem in the whole space V_0, we apply a penalty method [3, 7, 9, 13] to problem (12). We use a penalty in the form

$$J_{\theta}(\mathbf{u}) = \frac{1}{2\theta}\sum_{\{\alpha,\,\beta\}\in Q}\left\|\left(d_{\alpha\beta} - u_{\alpha n} - u_{\beta n}\right)^{-}\right\|^2_{L_2(S_{\alpha\beta})} + \\ + \frac{1}{2\theta}\sum_{\{\alpha',\,\beta'\}\in Q'}\left\|\mathbf{u}_{\alpha'} - \mathbf{u}_{\beta'}\right\|^2_{[L_2(\Gamma_{\alpha'\beta'})]^3}, \tag{20}$$

where $\theta > 0$ is a penalty parameter, and $y^{-} = \min\{0, y\}$.

Now, consider the following unconstrained minimization problem in V_0:

$$F_{\theta}(\mathbf{u}) = A(\mathbf{u},\mathbf{u})/2 - H(\mathbf{u}) - L(\mathbf{u}) + J_{\theta}(\mathbf{u}) \to \min_{\mathbf{u}\in V_0}. \tag{21}$$

The penalty term J_{θ} is nonnegative and Gateaux differentiable in V_0, and its differential $J'_{\theta}(\mathbf{u},\mathbf{v}) = -\frac{1}{\theta}\sum_{\{\alpha,\,\beta\}\in Q}\int_{S_{\alpha\beta}}\left(d_{\alpha\beta} - u_{\alpha n} - u_{\beta n}\right)^{-}\left(v_{\alpha n} + v_{\beta n}\right)dS + \frac{1}{\theta}\sum_{\{\alpha',\,\beta'\}\in Q'}\int_{\Gamma_{\alpha'\beta'}}\left(\mathbf{u}_{\alpha'} - \mathbf{u}_{\beta'}\right)\cdot\left(\mathbf{v}_{\alpha'} - \mathbf{v}_{\beta'}\right)dS$ satisfies the following properties [15]:

$$(\forall \mathbf{u}\in V_0)(\exists \tilde{R}>0)(\forall \mathbf{v}\in V_0)\left\{\left|J'_{\theta}(\mathbf{u},\mathbf{v})\right| \leq \tilde{R}\,\|\mathbf{v}\|_{V_0}\right\}, \tag{22}$$

$$(\exists \tilde{D}>0)(\forall \mathbf{u},\mathbf{v},\mathbf{w}\in V_0)\left\{\left|J'_{\theta}(\mathbf{u}+\mathbf{w},\mathbf{v}) - J'_{\theta}(\mathbf{u},\mathbf{v})\right| \leq \tilde{D}\,\|\mathbf{v}\|_{V_0}\|\mathbf{w}\|_{V_0}\right\}, \tag{23}$$

$$(\forall \mathbf{u},\mathbf{v}\in V_0)\left\{J'_{\theta}(\mathbf{u}+\mathbf{v},\mathbf{v}) - J'_{\theta}(\mathbf{u},\mathbf{v}) \geq 0\right\}. \tag{24}$$

Using these properties and the results in [3], we have shown that problem (21) has a unique solution $\bar{\mathbf{u}}_{\theta} \in V_0$ and is equivalent to the following nonlinear variational equation in the space V_0:

$$F'_{\theta}(\mathbf{u},\mathbf{v}) = A(\mathbf{u},\mathbf{v}) - H'(\mathbf{u},\mathbf{v}) + J'_{\theta}(\mathbf{u},\mathbf{v}) - L(\mathbf{v}) = 0,\ \forall \mathbf{v}\in V_0,\ \mathbf{u}\in V_0. \tag{25}$$

Using the results of works [7, 13], we have proved that $\|\bar{\mathbf{u}}_{\theta} - \bar{\mathbf{u}}\|_{V_0} \underset{\theta\to 0}{\to} 0$.

4 Iterative Methods for Nonlinear Variational Equations

In arbitrary reflexive Banach space V_0 consider an abstract nonlinear variational equation

$$\Phi(\mathbf{u},\mathbf{v}) = L(\mathbf{v}),\ \forall \mathbf{v} \in V_0,\ \mathbf{u} \in V_0, \tag{26}$$

where $\Phi : V_0 \times V_0 \to \mathbb{R}$ is a functional, which is linear in $\mathbf{v}$, but nonlinear in $\mathbf{u}$, and L is linear continuous form. Suppose that this variational equation has a unique solution $\bar{\mathbf{u}}_* \in V_0$.

For the numerical solution of (26) we use the next iterative method [5, 6, 15]:

$$G(\mathbf{u}^{k+1},\mathbf{v}) = G(\mathbf{u}^k,\mathbf{v}) - \gamma\left[\Phi(\mathbf{u}^k,\mathbf{v}) - L(\mathbf{v})\right],\ \forall \mathbf{v} \in V_0,\ k = 0,1,\dots, \tag{27}$$

where G is some given bilinear form in V_0, $\gamma \in \mathbb{R}$ is fixed parameter, and $\mathbf{u}^k \in V_0$ is the k-th approximation to the exact solution of problem (26).

We have proved the next theorem [5, 15] about the convergence of this method.

Theorem 1. *Suppose that the following conditions hold*

$$(\forall \mathbf{u} \in V_0)\,(\exists R_\Phi > 0)\,(\forall \mathbf{v} \in V_0)\left\{\,|\Phi(\mathbf{u},\mathbf{v})| \le R_\Phi \|\mathbf{v}\|_{V_0}\right\}, \tag{28}$$

$$(\exists D_\Phi > 0)(\forall \mathbf{u},\mathbf{v},\mathbf{w} \in V_0)\left\{|\Phi(\mathbf{u}+\mathbf{w},\mathbf{v}) - \Phi(\mathbf{u},\mathbf{v})| \le D_\Phi \|\mathbf{v}\|_{V_0}\|\mathbf{w}\|_{V_0}\right\}, \tag{29}$$

$$(\exists B_\Phi > 0)\,(\forall \mathbf{u},\mathbf{v} \in V_0)\left\{\,\Phi(\mathbf{u}+\mathbf{v},\mathbf{v}) - \Phi(\mathbf{u},\mathbf{v}) \ge B_\Phi \|\mathbf{v}\|^2_{V_0}\right\}, \tag{30}$$

bilinear form G is symmetric, continuous with constant $M_G > 0$ and coercive with constant $B_G > 0$, and $\gamma \in (0; 2\gamma^*), \gamma^* = B_\Phi B_G / D^2_\Phi$.

Then $\left\|\mathbf{u}^k - \bar{\mathbf{u}}_*\right\|_{V_0} \underset{k\to\infty}{\to} 0$, where $\{\mathbf{u}^k\} \subset V_0$ is obtained by method (27). Moreover, the convergence rate in norm $\|\cdot\|_G = \sqrt{G(\cdot,\cdot)}$ is linear, and the highest convergence rate in this norm reaches as $\gamma = \gamma^*$.

In addition, we have proposed nonstationary iterative method to solve (26), where bilinear form G and parameter γ are different in each iteration:

$$G^k(\mathbf{u}^{k+1},\mathbf{v}) = G^k(\mathbf{u}^k,\mathbf{v}) - \gamma^k\left[\Phi(\mathbf{u}^k,\mathbf{v}) - L(\mathbf{v})\right],\ \forall \mathbf{v} \in V_0,\ k = 0,1,\dots. \tag{31}$$

A convergence theorem for this method is proved in [15].

5 Domain Decomposition Schemes for Contact Problems

Now let us apply iterative methods (27) and (31) to the solution of nonlinear penalty variational equation (25) of multibody contact problem. This penalty equation can be written in form (26), where

$$\Phi(\mathbf{u},\mathbf{v}) = A(\mathbf{u},\mathbf{v}) - H'(\mathbf{u},\mathbf{v}) + J'_\theta(\mathbf{u},\mathbf{v}),\ \ \mathbf{u},\mathbf{v} \in V_0. \tag{32}$$

We propose such variants of methods (27) and (31), which lead to the domain decomposition.

Let us take the bilinear form G in iterative method (27) as follows [6, 15]:

$$G(\mathbf{u},\mathbf{v}) = A(\mathbf{u},\mathbf{v}) + X(\mathbf{u},\mathbf{v}),\ \ \mathbf{u},\mathbf{v} \in V_0, \tag{33}$$

$$X(\mathbf{u},\mathbf{v}) = \frac{1}{\theta}\sum_{\alpha=1}^{N}\left[\sum_{\beta\in B_\alpha}\int_{S_{\alpha\beta}} u_{\alpha n} v_{\alpha n}\,\psi_{\alpha\beta}\,dS + \sum_{\beta'\in I_\alpha}\int_{\Gamma_{\alpha\beta'}} \mathbf{u}_\alpha\cdot\mathbf{v}_\alpha\,\phi_{\alpha\beta'}\,dS\right],$$

where $\psi_{\alpha\beta}(\mathbf{x}) = \{1, \mathbf{x}\in S^1_{\alpha\beta}\}\vee\{0,\ \mathbf{x}\in S_{\alpha\beta}\backslash S^1_{\alpha\beta}\}$ and $\phi_{\alpha\beta'}(\mathbf{x}) = \{1, \mathbf{x}\in \Gamma^1_{\alpha\beta'}\}\vee\{0,\ \mathbf{x}\in \Gamma_{\alpha\beta'}\backslash\Gamma^1_{\alpha\beta'}\}$ are characteristic functions of arbitrary subsets $S^1_{\alpha\beta}\subseteq S_{\alpha\beta}$, $\Gamma^1_{\alpha\beta'}\subseteq\Gamma_{\alpha\beta'}$ of possible unilateral and ideal contact areas respectively.

Introduce a notation $\tilde{\mathbf{u}}^{k+1} = [\mathbf{u}^{k+1}-\mathbf{u}^k]/\gamma + \mathbf{u}^k \in V_0$. Then iterative method (27) with bilinear form (33) can be written in such way:

$$A\left(\tilde{\mathbf{u}}^{k+1},\mathbf{v}\right) + X\left(\tilde{\mathbf{u}}^{k+1},\mathbf{v}\right) = L(\mathbf{v}) + X\left(\mathbf{u}^k,\mathbf{v}\right) + H'(\mathbf{u}^k,\mathbf{v}) - J'_\theta(\mathbf{u}^k,\mathbf{v}), \tag{34}$$

$$\mathbf{u}^{k+1} = \gamma\,\tilde{\mathbf{u}}^{k+1} + (1-\gamma)\,\mathbf{u}^k,\ \ k = 0,1,\dots\,. \tag{35}$$

Bilinear form X is symmetric, continuous with constant $M_X > 0$, and nonnegative [15]. Due to these properties, and due to the properties of bilinear form A, it follows that the conditions of Theorem 1 hold. Therefore, we obtain the next proposition:

Theorem 2. *The sequence $\{\mathbf{u}^k\}$ of the method (34)–(35) converges strongly to the solution of penalty variational equation (25) for $\gamma\in(0;\,2B_\Phi B_G/D^2_\Phi)$, where $B_G = B_A$, $B_\Phi = B$, $D_\Phi = M_A + D + \tilde{D}$. The convergence rate in norm $\|\cdot\|_G$ is linear.*

As the common quantities of the subdomains are known from the previous iteration, variational equation (34) splits into N separate equations for each subdomain Ω_α, and method (34)–(35) can be written in the following equivalent form:

$$a_\alpha(\tilde{\mathbf{u}}^{k+1}_\alpha,\mathbf{v}_\alpha) + \sum_{\beta\in B_\alpha}\int_{S_{\alpha\beta}}\frac{\psi_{\alpha\beta}}{\theta}\,\tilde{u}^{k+1}_{\alpha n}\,v_{\alpha n}\,dS + \sum_{\beta'\in I_\alpha}\int_{\Gamma_{\alpha\beta'}}\frac{\phi_{\alpha\beta'}}{\theta}\,\tilde{\mathbf{u}}^{k+1}_\alpha\cdot\mathbf{v}_\alpha\,dS$$

$$= l_\alpha(\mathbf{v}_\alpha) + \frac{1}{\theta}\sum_{\beta\in B_\alpha}\int_{S_{\alpha\beta}}\left[\psi_{\alpha\beta}\,u^k_{\alpha n} + \left(d_{\alpha\beta} - u^k_{\alpha n} - u^k_{\beta n}\right)^-\right]v_{\alpha n}\,dS$$

$$+\frac{1}{\theta}\sum_{\beta'\in I_\alpha}\int_{\Gamma_{\alpha\beta'}}\left[\phi_{\alpha\beta'}\,\mathbf{u}^k_\alpha + \left(\mathbf{u}^k_{\beta'} - \mathbf{u}^k_\alpha\right)\right]\cdot\mathbf{v}_\alpha\,dS + h'_\alpha(\mathbf{u}^k_\alpha,\mathbf{v}_\alpha),\ \ \forall\mathbf{v}_\alpha\in V^0_\alpha, \tag{36}$$

$$\mathbf{u}^{k+1}_\alpha = \gamma\,\tilde{\mathbf{u}}^{k+1}_\alpha + (1-\gamma)\,\mathbf{u}^k_\alpha,\ \ \alpha = 1,2,\dots,N,\ \ k = 0,1,\dots\,. \tag{37}$$

In each iteration k of method (36)–(37) we have to solve N linear variational equations in parallel, which correspond to some linear elasticity problems in subdomains with additional volume forces in Ω_α and with Robin boundary conditions on contact areas. Therefore, this method refers to parallel Robin–Robin type domain decomposition schemes.

Taking different characteristic functions $\psi_{\alpha\beta}$ and $\phi_{\alpha'\beta'}$, we can obtain different particular cases of penalty domain decomposition method (36)–(37).

Thus, taking $\psi_{\alpha\beta}(\mathbf{x}) \equiv 0$, $\beta \in B_\alpha$, $\phi_{\alpha\beta'}(\mathbf{x}) \equiv 0$, $\beta' \in I_\alpha$, $\alpha = 1,2,\dots,N$, we get parallel Neumann–Neumann domain decomposition scheme.

Other borderline case is when $\psi_{\alpha\beta}(\mathbf{x}) \equiv 1$, $\beta \in B_\alpha$, $\phi_{\alpha\beta'}(\mathbf{x}) \equiv 1$, $\beta' \in I_\alpha$, $\alpha = 1,2,\dots,N$, i.e. $S^1_{\alpha\beta} = S_{\alpha\beta}$, $\Gamma^1_{\alpha\beta'} = \Gamma_{\alpha\beta'}$.

Moreover, we can choose functions $\psi_{\alpha\beta}$ and $\phi_{\alpha\beta'}$ differently in each iteration k. Then we obtain nonstationary domain decomposition schemes, which are equivalent to iterative method (31) with bilinear forms

$$G^k(\mathbf{u},\mathbf{v}) = A(\mathbf{u},\mathbf{v}) + X^k(\mathbf{u},\mathbf{v}), \quad \mathbf{u},\mathbf{v} \in V_0, \quad k = 0,1,\dots, \tag{38}$$

$$X^k(\mathbf{u},\mathbf{v}) = \frac{1}{\theta}\sum_{\alpha=1}^{N}\left[\sum_{\beta\in B_\alpha}\int_{S_{\alpha\beta}} u_{\alpha n} v_{\alpha n}\,\psi^k_{\alpha\beta}\,dS + \sum_{\beta'\in I_\alpha}\int_{\Gamma_{\alpha\beta'}} \mathbf{u}_\alpha\cdot\mathbf{v}_\alpha\,\phi^k_{\alpha\beta'}\,dS\right].$$

If we take characteristic functions $\psi^k_{\alpha\beta}$ and $\phi^k_{\alpha\beta'}$ as follows [6, 14, 15]:

$$\psi^k_{\alpha\beta}(\mathbf{x}) = \chi^k_{\alpha\beta}(\mathbf{x}) = \begin{cases} 0, & d_{\alpha\beta}(\mathbf{x}) - u^k_{\alpha n}(\mathbf{x}) - u^k_{\beta n}(\mathbf{x}') \ge 0 \\ 1, & d_{\alpha\beta}(\mathbf{x}) - u^k_{\alpha n}(\mathbf{x}) - u^k_{\beta n}(\mathbf{x}') < 0 \end{cases}, \ \mathbf{x}' = P(\mathbf{x}),\ \mathbf{x} \in S_{\alpha\beta},$$

$$\phi^k_{\alpha\beta'}(\mathbf{x}) \equiv 1, \ \mathbf{x} \in \Gamma_{\alpha\beta'}, \ \beta \in B_\alpha, \ \beta' \in I_\alpha, \ \alpha = 1,2,\dots,N,$$

then we shall get the method, which can be conventionally named as nonstationary parallel Dirichlet–Dirichlet domain decomposition scheme.

In addition to methods (27), (33) and (31), (38), we have proposed another family of DDMs for the solution of (25), where the second derivative of functional $H(\mathbf{u})$ is used. These domain decomposition methods are obtained from (31), if we choose bilinear forms $G^k(\mathbf{u},\mathbf{v})$ as follows

$$G^k(\mathbf{u},\mathbf{v}) = A(\mathbf{u},\mathbf{v}) - H''(\mathbf{u}^k,\mathbf{u},\mathbf{v}) + X^k(\mathbf{u},\mathbf{v}), \quad \mathbf{u},\mathbf{v} \in V_0, \quad k = 0,1,\dots. \tag{39}$$

Numerical analysis of presented penalty Robin–Robin DDMs has been made for plane unilateral two-body and three-body contact problems of linear elasticity ($\omega_\alpha \equiv 0$) using finite element approximations [6, 14, 15]. Numerical experiments have confirmed the theoretical results about the convergence of these methods.

Among the positive features of proposed domain decomposition schemes are the simplicity of the algorithms and the regularization of original contact problem because of the use of penalty method. These domain decomposition schemes have only one iteration loop, which deals with domain decomposition, nonlinearity of the stress-strain relationship, and nonlinearity of unilateral contact conditions.

Bibliography

[1] P. Avery and C. Farhat. The FETI family of domain decomposition methods for inequality-constrained quadratic programming: Application to contact problems with conforming and nonconforming interfaces. *Comput. Methods Appl. Mech. Engrg.*, 198:1673–1683, 2009.

[2] G. Bayada, J. Sabil, and T. Sassi. A Neumann–Neumann domain decomposition algorithm for the Signorini problem. *Appl. Math. Lett.*, 17(10):1153–1159, 2004.

[3] J. Céa. *Optimisation. Théorie et algorithmes*. Dunod, Paris, 1971. [In French].

[4] Z. Dostál, D. Horák, and D. Stefanica. A scalable FETI–DP algorithm with non-penetration mortar conditions on contact interface. *Journal of Computational and Applied Mathematics*, 231:577–591, 2009.

[5] I.I. Dyyak and I.I. Prokopyshyn. Convergence of the Neumann parallel scheme of the domain decomposition method for problems of frictionless contact between several elastic bodies. *Journal of Mathematical Sciences*, 171(4):516–533, 2010.

[6] I.I. Dyyak and I.I. Prokopyshyn. Domain decomposition schemes for frictionless multibody contact problems of elasticity. In G. Kreiss et al., editor, *Numerical Mathematics and Advanced Applications 2009*, pages 297–305. Springer Berlin Heidelberg, 2010.

[7] R. Glowinski, J.L. Lions, and R. Trémolières. *Analyse numérique des inéquations variationnelles*. Dunod, Paris, 1976. [In French].

[8] A.A. Ilyushin. *Plasticity*. Gostekhizdat, Moscow, 1948. [In Russian].

[9] N. Kikuchi and J.T. Oden. *Contact problem in elasticity: A study of variational inequalities and finite element methods*. SIAM, 1988.

[10] J. Koko. An optimization-bazed domain decomposition method for a two-body contact problem. *Num. Func. Anal. Optim.*, 24(5–6):586–605, 2003.

[11] R. Krause and B. Wohlmuth. A Dirichlet–Neumann type algorithm for contact problems with friction. *Comput. Visual. Sci.*, 5(3):139–148, 2002.

[12] A.S. Kravchuk. Formulation of the problem of contact between several deformable bodies as a nonlinear programming problem. *Journal of Applied Mathematics and Mechanics*, 42(3):489–498, 1978.

[13] J.L. Lions. *Quelques méthodes de résolution des problèmes aux limites non linéaire*. Dunod Gauthier-Villards, Paris, 1969. [In French].

[14] I.I. Prokopyshyn. Parallel domain decomposition schemes for frictionless contact problems of elasticity. *Visnyk Lviv Univ. Ser. Appl. Math. Comp. Sci.*, 14: 123–133, 2008. [In Ukrainian].

[15] I.I. Prokopyshyn. *Penalty method based domain decomposition schemes for contact problems of elastic solids*. PhD thesis, IAPMM NASU, Lviv, 2010. URL:194.44.15.230:8080/v25/resources/thesis/Prokopyshyn.pdf. [In Ukrainian].

[16] T. Sassi, M. Ipopa, and F.-X. Roux. Generalization of Lion's nonoverlapping domain decomposition method for contact problems. *Lect. Notes Comput. Sci. Eng.*, 60:623–630, 2008.

Domain Decomposition Method for Stokes Problem with Tresca Friction

Mohamed Khaled Gdoura[1], Jonas Koko[2], and Taoufik Sassi[3]

[1] LAMSIN, Université Tunis El Manar, BP 37, 1002 Tunis, TUNISIA;
LMNO, Université de Caen – CNRS UMR 6139
BP 5186, 14032 Caen Cedex, FRANCE; mohamedkhaled.gdoura@lamsin.rnu.tn
[2] LIMOS, Université Blaise-Pascal – CNRS UMR 6158
Campus des Cézeaux, 63173 Aubière cedex, FRANCE; koko@isima.fr
[3] LMNO, Université de Caen – CNRS UMR 6139
BP 5186, 14032 Caen Cedex, FRANCE; Taoufik.Sassi@math.unicaen.fr

1 Introduction

Development of numerical methods for the solution of Stokes system with slip boundary conditions (Tresca friction conditions) is a challenging task whose difficulty lies in the nonlinear conditions. Such boundary conditions have to be taken into account in many situations arising in practice, in flow of polymers (see [10] and references therein).

The paper is devoted to domain decomposition methods (DDM in short) for the Stokes problem with the slip boundary conditions. The original domain is cut into two sub-domains and the augmented Lagrangian formulation for separate resulting Poisson problems in both domains is used for computations. To relate solutions of these two sub-problems to the original solution, one has to introduce additional constraints "gluing" them together. The domain decomposition formulation is based on the Uzawa block relaxation method for the augmented Lagrangian involving three supplementary conditions. The paper is concluded by preliminary several numerical examples.

2 Setting Stokes Problem with Nonlinear Boundary Conditions

Let us consider a domain $\Omega \subset \mathbb{R}^2$ with the Lipschitz boundary $\partial\Omega$ which is split into two non-empty and non-overlapping parts Γ_0 and Γ. We denote by n the outward unit normal to $\partial\Omega$ and u_n, respectively u_t, the normal, respectively the tangential, component of u. We also make use of σ_t for the tangential component of the stress vector $\sigma(u)n$. The problem consists in finding the velocity field u and the pressure p for the following Stokes problem with nonlinear boundary condition of Tresca friction type:

R. Bank et al. (eds.), *Domain Decomposition Methods in Science and Engineering XX*, 655
Lecture Notes in Computational Science and Engineering 91,
DOI 10.1007/978-3-642-35275-1_78, © Springer-Verlag Berlin Heidelberg 2013

$$\begin{cases} -div(\nu\varepsilon(u)) + \nabla p = \mathrm{f} & \text{in } \Omega \\ div(u) = 0 & \text{in } \Omega \\ u = 0 & \text{on } \Gamma_0 \\ u_n = 0 & \text{on } \Gamma \\ |\sigma_t| \le g & \text{on } \partial\Omega \\ |\sigma_t| < g \Rightarrow \quad u_t = 0 & \text{on } \Gamma \\ |\sigma_t| = g \Rightarrow \exists k > 0 \text{ a constant such that } \quad u_t = -k\sigma_t & \text{on } \Gamma \end{cases} \tag{1}$$

where f is in $\mathbf{L}^2(\Omega)$, $g \in L^2(\Gamma)$, $g > 0$ is the given slip bound on Γ and $|\cdot|$ is the euclidean norm.

One can derive the variational formulation of (1):

$$\begin{cases} \text{Find } u \in \mathbf{V}_{div}(\Omega) \text{ such that} : \forall\, v \in \mathbf{V}_{div}(\Omega) \\ a(u, v-u) + j(v) - j(u) \ge L(v-u), \end{cases} \tag{2}$$

with

$$\mathbf{V}(\Omega) = \left\{ v \in \mathbf{H}^1(\Omega),\ v_{|\Gamma_0} = 0, v_n = 0 \text{ on } \Gamma \right\},$$

$$\mathbf{V}_{div}(\Omega) = \left\{ v \in \mathbf{V}(\Omega)\ ,\ div(v) = 0 \text{ in } \Omega \right\},$$

$$a(u,v) = \int_\Omega \nu\varepsilon(u) : \varepsilon(v)\, d\Omega, \quad L(v) = \int_\Omega \mathrm{f} v\, d\Omega, \quad j(v) = \int_\Gamma g|v_t|\, d\Gamma.$$

Problem (2) is an elliptic variational inequality of the second kind which has a unique solution [3]. Moreover, since the bilinear form $a(\cdot,\cdot)$ is symmetric (2) is equivalent to the following constrained non-differentiable minimization problem:

$$\text{Find } u \in \mathbf{V}_{div}(\Omega) \text{ such that} : \quad \mathscr{J}(u) \le \mathscr{J}(v) \quad \forall v \in \mathbf{V}_{div}(\Omega), \tag{3}$$

where $\mathscr{J}(v) = \dfrac{1}{2} a(v,v) + j(v) - L(v)$ is the total potential energy functional.

3 Uzawa DDM for Stokes Problem with Tresca Friction

We now study the domain decomposition of (3). We first rewrite (3) in the following more useful form. Suppose that $\varphi = v_t$, then the minimization problem (3) becomes:

$$\begin{cases} \text{Find } (u, \Phi) \in \Pi \text{ such that:} \\ \\ \Sigma(u, \Phi) \le \Sigma(v, \varphi) \forall\, (v, \varphi) \in \Pi, \end{cases} \tag{4}$$

where

$$\Pi = \{(v, \varphi) \in \mathbf{V}_{div}(\Omega) \times H^{\frac{1}{2}}(\Gamma) \text{ such that } \varphi = v_t\},$$

and Σ is the Lagrangian defined on Π by:

$$\forall(\varphi, v) \in \Pi \qquad \Sigma(v, \varphi) = \frac{1}{2}a(v, v) - L(v) + j(\varphi). \tag{5}$$

Let $\{\Omega_1, \Omega_2\}$ be a partition of Ω, as shown in Fig. 1, and let

$$\Gamma_{12} = \Gamma_{21} = \partial\Omega_1 \cap \partial\Omega_2, \quad \Gamma_i = \Gamma \cup \partial\Omega_i, \quad \Gamma_i^0 = \Gamma_0 \cup \partial\Omega_i,$$
$$v_i = v|_{\Omega_i}, \quad p_i = p|_{\Omega_i},$$
$$\mathbf{V}(\Omega_i) = \left\{ v_i \in \mathbf{H}^1(\Omega_i),\ v_{i|\Gamma_i^0} = 0,\ v_i.n_{i|\Gamma_i} = 0 \right\},$$
$$\mathbf{V}_{div}(\Omega_i) = \left\{ v_i \in \mathbf{V}(\Omega_i)\ ,\ div(v_i) = 0 \text{ in } \Omega_i \right\}.$$

Restrictions of the functionals a and Σ over Ω_i are defined by a_i and Σ_i respectively. Inner products over a given part S of $\partial\Omega_i$, $i = 1, 2$, and Ω_i are defined by

$$(u, v)_S = \int_S uv\mathrm{d}\Gamma \quad \text{and} \quad (u, v)_{\Omega_i} = \int_{\Omega_i} uv\mathrm{d}x.$$

We treat the pressure as a Lagrange multiplier associated with the constraint $div(u) =$

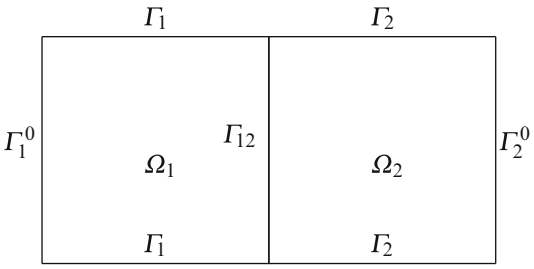


Fig. 1. Decomposition of Ω into two subdomains

0. Using the decomposition of Fig. 1, the functional (5) becomes

$$\Sigma(v, \varphi) = \Sigma_1(v_1, \varphi_1) + \Sigma_2(v_2, \varphi_2). \tag{6}$$

It is clear that problem (3) is equivalent to the following constrained minimization problem:

$$\begin{aligned} \forall(v_i, \varphi_i) \in \mathbf{V}(\Omega_i) \times H^{\frac{1}{2}}(\Gamma_i), i = 1, 2 \\ \Sigma(u_1, \Phi_1) + \Sigma(u_2, \Phi_2) \leq \Sigma_1(v_1, \varphi_1) + \Sigma_2(v_2, \varphi_2) \\ div(u_i) = 0 \quad \text{in } \Omega_i, \\ u_{it} - \Phi_i = 0 \quad \text{in } \Gamma_i, \\ u_i - \Psi = 0 \quad \text{in } \Gamma_{12}. \end{aligned} \tag{7}$$

The auxiliary interface unknown Ψ is added to the continuity constraint to avoid coupling between u_1 and u_2 in the penalty term. This so-called *three-field formulation* has been used in domain decomposition of elliptic problems [9]. To ensure the uniqueness of the pressure, the following constraint can be added

$$\int_{\Omega_1} p_1 \, d\Omega_1 + \int_{\Omega_2} p_2 \, d\Omega_1 = 0. \tag{8}$$

Then, we introduce the set

$$\mathfrak{P} = \left\{ (q_1, q_2) \in L^2(\Omega_1) \times L^2(\Omega_2) \text{ such that } \int_{\Omega_1} q_1 \, d\Omega_1 + \int_{\Omega_2} q_2 \, d\Omega_1 = 0 \right\}$$

We can associate to (7) the augmented Lagrangian functional $\mathscr{L}_r$ defined by

$$\begin{aligned} \mathscr{L}_r \ (u, \Phi, \Psi, p, \mu, \lambda) &= \Sigma(u_1, \Phi_1) + \Sigma(u_2, \Phi_2) \\ &+ \sum_{i=1}^{2} \left[(\mu_i, \Phi_i - u_{it})_{\Gamma_i} - (p_i, div(u_i))_{\Omega_i} + (\lambda_i, u_i - \Psi)_{\Gamma_{12}} \right] \\ &+ \sum_{i=1}^{2} \left[\frac{r_1}{2} ||div(u_i)||^2_{L^2(\Omega_i)} + \frac{r_2}{2} ||\Phi_i - u_{it}||^2_{L^2(\Gamma_i)} + \frac{r_3}{2} ||u_i - \Psi||^2_{L^2(\Gamma_{12})} \right]. \end{aligned} \tag{9}$$

where r_1, r_2 and r_3 are the penalty parameters which are strictly positive.

Remark 1. The standard L^2 scalar product (not equivalent to the $H^{1/2}$ scalar product) on the interface Γ_{12} and Γ_i is used in the definition of (9). This approach is easy to implement but it has some negative effects on the convergence of our algorithm.

Then, problem (7) is equivalent to the following saddle-point problem:

$$\begin{cases} \text{Find } (u, \Phi, \Psi, p, \mu, \lambda) \in \mathscr{H} \quad \text{such that: } \forall (v, \Phi, \Psi, q, \tilde{\mu}, \tilde{\lambda}) \in \mathscr{H} \\ \\ \mathscr{L}_r(u, \Phi, \Psi, q, \tilde{\mu}, \tilde{\lambda}) \leq \mathscr{L}_r(u, \Phi, \Psi, p, \mu, \lambda) \leq \mathscr{L}_r(v, \Phi, \Psi, p, \mu, \lambda). \end{cases} \tag{10}$$

where $u{=}(u_1, u_2){\in}\mathbf{V}(\Omega_1) \times \mathbf{V}(\Omega_2)$, $\Phi{=}(\Phi_1, \Phi_2) \in L^2(\Gamma_1) \times L^2(\Gamma_2)$, $\Psi \in (L^2(\Gamma_{12}))^2$, $p = (p_1, p_2) \in \mathfrak{P}$, $\mu = (\mu_1, \mu_2) \in L^2(\Gamma_1) \times L^2(\Gamma_2)$ and $\lambda \in (L^2(\Gamma_{12}))^2$. $\mathscr{H}$ is the Cartesian product of all these spaces.

3.1 Uzawa Block Relaxation Method: UBR2

In order to solve (10) we use Uzawa block relaxation algorithm based on ALG2, see [4]. This leads to the following iterations:

`Initialization:` Φ^{-1}, Ψ^{-1}, p^0, λ^0, μ^0 `and` $r_i > 0$ `fixed.`
`Repeat until convergence:`

`1. Find` $u^k \in \mathbf{V}(\Omega_1) \times \mathbf{V}(\Omega_2)$ `such that:` $\quad \forall v \in \mathbf{V}(\Omega_1) \times \mathbf{V}(\Omega_2)$

$$\mathscr{L}_r(u^k, \Phi^{k-1}, \Psi^{k-1}, p^k, \mu^k, \lambda^k) \leq \mathscr{L}_r(v, \Phi^{k-1}, \Psi^{k-1}, p^k, \mu^k, \lambda^k). \tag{11}$$

2. Find $\Phi^k \in L^2(\Gamma_1) \times L^2(\Gamma_2)$ such that: $\forall \Phi \in L^2(\Gamma_1) \times L^2(\Gamma_2)$

$$\mathscr{L}_r(u^k, \Phi^k, \Psi^{k-1}, p^k, \mu^k, \lambda^k) \leq \mathscr{L}_r(u^k, \Phi, \Psi^{k-1}, p^k, \mu^k, \lambda^k). \tag{12}$$

3. Find $\Psi^k \in (L^2(\Gamma_{12}))^2$ such that: $\forall \Psi \in (L^2(\Gamma_{12}))^2$.

$$\mathscr{L}_r(u^k, \Phi^k, \Psi^k, p^k, \mu^k, \lambda^k) \leq \mathscr{L}_r(u^k, \Phi^k, \Psi, p^k, \mu^k, \lambda^k). \tag{13}$$

4. Lagrange multipliers update

$$p_i^{k+1} = p_i^k - r_1 div(u_i^k), \tag{14}$$
$$\lambda_i^{k+1} = \lambda_i^k + r_2(u_{i|\Gamma_{12}}^k - \Psi^k), \tag{15}$$
$$\mu_i^{k+1} = \mu_i^k + r_3(u_{it}^k - \Phi_i^k). \tag{16}$$

Subproblem (11) is equivalent to solving, in each subdomain, the following problem:

Find $u_i^k \in \mathbf{V}(\Omega_i)$ such that

$$a(u_i^k, v) + r_1(\nabla.u_i^k, \nabla.v_i)_{\Omega_i} + r_2(u_i, v_i)_{\Gamma_{12}} + r_3(u_t^k, v_t)_{\Gamma} = (\mathrm{f}_i, v_i) + (p_i, \nabla.v_i)_{\Omega_i}$$
$$+ (r_2\Psi^k - \lambda^k, v_i)_{\Gamma_{12}} + (r_3\Phi_i^{k-1} - \mu_i^k, v_{it})_{\Gamma_i} \quad \forall v_i \in \mathbf{V}(\Omega_i). \tag{17}$$

The subproblems of steps 2 and 3 are uncoupled and consists in the following calculations:

$$\Phi_i^k = \begin{cases} \dfrac{||\mu_i^k + r_3 u_{it}^k||_{0,\Gamma_i} - g}{r_3||\mu_i^k + r_3 u_{it}^k||_{0,\Gamma_i}}(\mu_i^k + r_3 u_{it}^k) & \text{if} \quad ||\mu_i^k + r_3 u_{it}^k||_{0,\Gamma_i} \geq g \\ \\ 0 \text{ unless} \end{cases} \tag{18}$$

and

$$\Psi^k = \frac{1}{2r_2}(\lambda_1^k + \lambda_2^k) + \frac{1}{2}(u_1^k + u_2^k)|_{\Gamma_{12}}. \tag{19}$$

Remark 2. For sake of simplicity the given slip bound g is assumed to be non-negative constant in (18).

Remark 3. After update (14), p^{k+1} must be projected onto $\mathfrak{P}$ to ensure the uniqueness of the pressure.

Remark 4. The main advantage of this formulation is that (17) reduces to 2D uncoupled elliptic problems which can be solved in parallel. Moreover, the matrices derived from discret problems are symmetric and positive definite.

4 Numerical Experiments

The domain decomposition algorithm **UBR2**, with $r_1 = r_2 = r_3$, presented in the previous section was implemented in Matlab V7.9 on a Core2 Duo-1.8 Ghz processor PC. For discrete velocity-pressure-Lagrange multipliers spaces, we use the P^1-iso-P^2/P^1 finite element. These spaces are well known to satisfy the discrete Babuska-Brezzi inf-sup condition [1].
For all the numerical experiments presented, the domain Ω is the square $[0,0.1]^2$, while $\Omega_1 = [0,0.05] \times [0,0.1]$ and $\Omega_2 = [0.05,0.1] \times [0,0.1]$. The fluid can slip on $\Gamma_1 \cup \Gamma_2 = [0,0.1] \times \{0.1\} \cup [0,0.1] \times \{0\}$, We set $g = 0.015$ which is consistent with experimental values, see [5]. The viscosity is taken equal to 0.1 and the stopping tolerance ε is 10^{-6}. In addition we enforce parabolic profile on both $\Gamma_1^0 = \{0\} \times [0,0.1]$ and $\Gamma_2^0 = \{0.1\} \times [0,0.1]$:

$$u|_{\Gamma_1^0} = u|_{\Gamma_2^0} = \begin{bmatrix} y(1-y) \\ -y(1-y) \end{bmatrix}$$

Remark 5. We choose this profile to enforce shear stress near the solid wall to reach the threshold without considering a complicated domain geometry.

In Fig. 2 we report the velocity field for the solution of Stokes problem with Tresca friction (1) in Ω and in $\Omega_1 \cup \Omega_2$. We can see that we have the same velocity profile. In Table 1 we report the discrete mesh size h, the corresponding number of degree

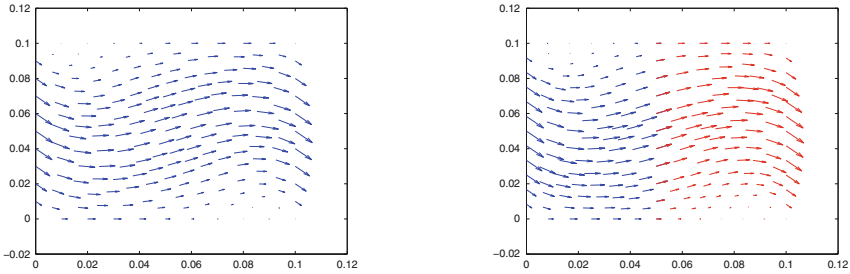


Fig. 2. Fluid flow with Tresca BC for one (*left*) and two domains (*right*)

of freedom (d.o.f) and number of elements on each subdomain in the follows experiments. Table 2 shows the number of iterations IT, the sequential CPU (in seconds) times and the parallel CPU* times (when subproblems (17) for $i = 1,2$ are solved in parallel). For several mesh size and for $\mathsf{N_{SD}}$ (Number of Sub-Domains) equal to 1 or 2. We notice that the **UBR2** algorithm is a h-dependent algorithm and the domain decomposition method to be preferable when dealing with parallel computing using parallel solver.

Table 3 show how the number of iterations and the optimal value of the relaxation parameter r_{opt} depend on h. We remark that the speed of convergence is very sensitive to r; this explains the strong increase in the number of iterations for a finer mesh.

N_{SD}	$h = 0.02$	$h = 0.01$	$h = 0.0067$	$h = 0.005$	$h = 0,004$
	$n/n_{\triangle}$	$n/n_{\triangle}$	$n/n_{\triangle}$	$n/n_{\triangle}$	$n/n_{\triangle}$
1	189/336	665/1284	1577/3032	2829/5496	4393/8548
2	112/188	370/676	806/1516	1396/2668	2220/4284

Table 1. h: mesh size; n: number of d.o.f. by domain $n_{\triangle}$: number of elements by domain.

N_{SD}	$h = 0.02$	$h = 0.01$	$h = 0.0067$	$h = 0.005$	$h = 0,004$
	IT/CPU/CPU*	IT/CPU/CPU*	IT/CPU/CPU*	IT/CPU/CPU*	IT/ CPU/CPU*
1	199/0.41/-	349/2.8/-	453/10.8/-	509/30.36/-	595/67.3/-
2	486/1/0.81	769/4.8/3.27	993/15.3/7.96	1294/41.14/21.98	1599/99.34/51.59

Table 2. Standard speed-up for h: mesh size; IT: number of iterations; CPU & CPU*: CPU times.

N_{SD}	$h = 0.02$	$h = 0.01$	$h = 0.0067$	$h = 0.005$	$h = 0,004$
	r_{opt}/IT	r_{opt}/IT	r_{opt}/IT	r_{opt}/IT	r_{opt}/IT
1	335/199	590/349	740/453	840/509	1010/595
2	116/486	124/769	175/993	230/1294	290/1599

Table 3. Convergence rate with respect r_{opt}.

5 Conclusion

The augmented Lagrangian formulation (9) of domain decomposed Stokes problem with Tresca friction leads to a numerical strategy which solves a classical Poisson problem (17) (in each subdomain Ω_i) and the contribution of Tresca friction (18) in a decoupled way. Nevertheless, this algorithm has a mesh dependent convergence and its practical implementation still facing the issue of the optimal choice of the penalties, $r_i, i = 1,2,3$. To improve this algorithm, different preconditioners will be investigated, especially the Steklov-Poincaré operator on the interface (see e.g. [6–8]) and the Cahouet-Chabard preconditioner [2] for the pressure multiplier.

Bibliography

[1] GDOURA M. K. AYADI M. and SASSI T. Mixed formulation for stokes problem with tresca friction. *C. R. Acad. Sci. Paris*, Ser. I 348:1069–1072, 2010.

[2] CAHOUET J. and CHABARD J. P. Some fast 3-D solvers for the generalized Stokes problem. *Internat. J. Numer. Methods Fluids*, 8:269–295, 1988.

[3] FUJITA H. A coherent analysis of stokes flows under boundary conditions of friction type. *J. Comput. and Appl. Math.*, 149:57–69, 2002.

[4] GLOWINSKI R. and LE TALLEC P. *Augmented Lagrangian and Operator-splitting Methods in Nonlinear Mechanics*. SIAM, 1989.

[5] HERVET H. and LÉGER M. Folw with slip at the wall: from simple to complex fluids. *C.R. Physique*, 4:241–249, 2003.

[6] KOKO J. Convergence analysis of optimization-based domain decomposition methods for a bonded structure. *Appl. Numer. Math.*, 58:69–87, 2008.

[7] KOKO J. and SASSI T. An uzawa domain decomposition method for stokes problem. *Domain Decomposition Methods in Science and Engineering XIX, Lecture Notes in Computational Science and Engineering*, 79:383–390, 2011.

[8] LE TALLEC P. and SASSI T. Domain decomposition with nonmatching grids: augmented Lagrangian approach. *Math. Comp.*, 64:1367–1396, 1995.

[9] QUARTERONI A. and VALLI A. *Domain Decomposition methods for partial differential equations*. Oxford University Press, 1999.

[10] RAO I. J. and RAJAGOPAL K.R. The effect of the slip boundary condition on the flow of fluids in a channel. *Acta Mechanica*, 135:113–126, 1999.

A Hybrid Discontinuous Galerkin Method for Darcy-Stokes Problems

Herbert Egger[1] and Christian Waluga[2]

[1] Technische Universität Darmstadt, Department of Mathematics, Numerical Analysis and Scientific Computing Dolivostraße 15, 64293 Darmstadt, Germany
egger@mathematik.tu-darmstadt.de

[2] Center of Mathematics, Technische Universität München Boltzmannstraße 3, 85748 Garching bei München, Germany
christian.waluga@ma.tum.de

Summary. We propose and analyze a hybrid discontinuous Galerkin method for the solution of incompressible flow problems, which allows to deal with pure Stokes, pure Darcy, and coupled Darcy-Stokes flow in a unified manner. The flexibility of the method is demonstrated in numerical examples.

1 Model Problem

Let $\Omega \subset \mathbb{R}^d$ be a bounded Lipschitz domain in $d = 2$ or 3 dimensions. Given data $\mathbf{f} \in [L^2(\Omega)]^d$ and $g \in L^2(\Omega)$, we consider the generalized Stokes problem

$$\sigma \mathbf{u} - 2\mu \operatorname{div} \varepsilon(\mathbf{u}) + \nabla p = \mathbf{f} \qquad \text{and} \qquad \operatorname{div} \mathbf{u} = g \quad \text{in } \Omega. \tag{1}$$

As usual, $\mathbf{u}$ denotes the velocity, p the pressure, and $\varepsilon(\mathbf{u}) := \frac{1}{2}(\nabla \mathbf{u} + \nabla \mathbf{u}^T)$ is the symmetric part of the velocity gradient tensor. We require that

$$\sigma \geq 0, \quad \mu \geq 0, \quad \text{and} \quad M \geq \sigma + \mu \geq m > 0 \qquad \text{in } \Omega.$$

For convenience, we assume that σ, the reciprocal of the permeability, and the viscosity μ are constant, and consider homogeneous boundary conditions

$$\mathbf{u}|_{\partial\Omega} = 0 \quad \text{if} \quad \mu > 0 \qquad \text{or} \qquad \mathbf{u} \cdot \mathbf{n}|_{\partial\Omega} = 0 \quad \text{if} \quad \mu = 0. \tag{2}$$

The unique solvability of the boundary value problem (1)–(2) is guaranteed, if the pressure p and the data g have zero average. For the case $\mu > 0$, we then have $(\mathbf{u}, p) \in \mathbf{H}^1_0(\Omega) \times L^2_0(\Omega)$, where $\mathbf{H}^1_0(\Omega) := \{\mathbf{v} \in [H^1(\Omega)]^d : \mathbf{v}|_{\partial\Omega} = 0\}$ and $L^2_0 := \{q \in L^2(\Omega) : \int_\Omega q \, dx = 0\}$. In the Darcy limit $\mu = 0$, we only have $\mathbf{u} \in \mathbf{H}_0(\operatorname{div}; \Omega) := \{\mathbf{v} \in [L^2(\Omega)]^d : \operatorname{div} \mathbf{v} \in L^2(\Omega),\ \mathbf{v} \cdot \mathbf{n}|_{\partial\Omega} = 0\}$.

For the approximation of problem (1)–(2), we consider a hybrid discontinuous Galerkin method, which is capable of treating incompressible flow in the Stokes

R. Bank et al. (eds.), *Domain Decomposition Methods in Science and Engineering XX*, Lecture Notes in Computational Science and Engineering 91,
DOI 10.1007/978-3-642-35275-1_79,

and Darcy regimes, as well as coupled problems in a unified manner. Our analysis extends the results of [7] for Stokes flow. Related work on stabilized non-conforming and discontinuous Galerkin methods for Darcy-Stokes flow can be found in [4, 8] and the references given there. We refer to [1, 5] for a unified treatment of discontinuous Galerkin methods for elliptic problems and their hybridization.

2 Notation and Preliminaries

Let $\mathscr{T}_h = \{T\}$ be a shape-regular quasi-uniform partition of Ω into affine families of triangles and/or quadrilaterals (tetrahedra and/or hexahedra) of size $\mathtt{h}$. By $\partial\mathscr{T}_h := \{\partial T : T \in \mathscr{T}_h\}$, we denote the set of element boundaries, and by $\mathscr{E}_h := \{E_{ij} = \partial T_i \cap \partial T_j : i > j\} \cup \{E_{i,0} = \partial T_i \cap \partial\Omega\}$ the set of edges (faces) between elements or on the boundary; $\mathscr{E} = \bigcup_{E\in\mathscr{E}_h} E$ is called the *skeleton*.

For $s \geq 0$, let $H^s(\mathscr{T}_h) := \{v \in L^2(\Omega) : v|_T \in H^s(T) \text{ for all } T \in \mathscr{T}_h\}$ denote the broken Sobolev space with inner product $(u,v)_{s,\mathscr{T}_h} := \sum_{T\in\mathscr{T}_h}(u,v)_{s,T}$ and norm $\|u\|_{s,\mathscr{T}_h}$; the subindex is omitted for $s = 0$. Piecewise defined derivatives are denoted with the standard symbols. The traces of functions in $H^1(\mathscr{T}_h)$ lie in $L^2(\partial\mathscr{T}_h)$, which is equipped with the scalar product $\langle u,v\rangle_{\partial\mathscr{T}_h} := \sum_{T\in\mathscr{T}_h}\langle u,v\rangle_{\partial T}$ and norm $|v|_{\partial\mathscr{T}_h}$. Any function in $L^2(\mathscr{E})$ can be identified with a function in $L^2(\partial\mathscr{T}_h)$ by doubling its values on the element interfaces. Bold symbols are used for vector valued functions.

Let $\mathscr{P}_{\mathtt{p}}(T)$ denote the polynomials of degree $\leq \mathtt{p}$ over T, and recall that

$$|v_{\mathtt{p}}|^2_{\partial T} \leq c_T\,\mathtt{p}^2\mathtt{h}^{-1}\|v_{\mathtt{p}}\|^2_T \qquad \text{for all } v_{\mathtt{p}} \in \mathscr{P}_{\mathtt{p}}(T). \tag{3}$$

Explicit bounds for the constant c_T in the discrete trace inequaliy (3) are known for all elements under consideration. The parameter c_T can be replaced by the shape regularity parameter $\gamma := \max\{c_T : T \in \mathscr{T}_h\}$, which is assumed to be independent of $\mathtt{h}$. We then choose a stabilization parameter α such that

$$4\gamma\mathtt{p}^2\mathtt{h}^{-1} \leq \alpha \leq 4\gamma'\mathtt{p}^2\mathtt{h}^{-1}, \tag{4}$$

with γ' independent of $\mathtt{p}$ and $\mathtt{h}$, and we define two norms on $L^2(\partial\mathscr{T}_h)$ by

$$|v|_{\pm1/2,\partial\mathscr{T}_h} := \big(\sum\nolimits_{T\in\mathscr{T}_h}|v|^2_{\pm1/2,\partial T}\big)^{1/2} \quad \text{with} \quad |v|_{\pm1/2,\partial T} := \alpha^{\pm1/2}|v|_{\partial T}.$$

Similar norms are frequently used for the analysis of mixed, non-conforming and discontinuous Galerkin methods; see e.g. [1].

3 The Hybrid DG Method

Let us fix $\mathtt{p} \geq 1$, and choose $\mathtt{q} = \mathtt{p} - 1$ or $\mathtt{q} = \mathtt{p}$. For the approximation of velocity and pressure in (1)–(2), we will utilize the finite element spaces

$$\mathbf{V}_h := \{\mathbf{v}_h \in \mathbf{L}^2(\mathscr{T}_h) : \mathbf{v}_h|_T \in [\mathscr{P}_{\mathtt{p}}(T)]^d \quad \text{for all } T \in \mathscr{T}_h\},$$
$$Q_h := \{q_h \in L^2_0(\Omega) : q_h|_T \in \mathscr{P}_{\mathtt{q}}(T) \quad \text{for all } T \in \mathscr{T}_h\}.$$

We further choose $\hat{\mathtt{p}} = \mathtt{p}$ or $\hat{\mathtt{p}} = \mathtt{q}$, and define a space

$$\widehat{\mathbf{V}}_h := \{\widehat{\mathbf{v}}_h \in \mathbf{L}^2(\mathscr{E}) : \widehat{\mathbf{v}}_h|_E \in [\mathscr{P}_{\hat{\mathtt{p}}}(E)]^d \quad \text{for all } E \in \mathscr{E}_h, \ \ \widehat{\mathbf{v}}_h = 0 \text{ on } \partial\Omega\},$$

of piecewise polynomials for representing velocities on the skeleton. The conditions $\mathtt{p} - 1 \le \mathtt{q} \le \mathtt{p}$ and $\mathtt{q} \le \hat{\mathtt{p}}$ are explicitly used in the analysis of a Fortin operator; see Proposition 5. In view of Lemma 1, we also require that $\hat{\mathtt{p}} \ge 1$. Note that the Dirichlet boundary condition has been included explicitly in the definition of the hybrid space $\widehat{\mathbf{V}}_h$. We further denote by $\pi^{\mathtt{p}} : \mathbf{H}^1(\mathscr{T}_h) \to \mathbf{V}_h$ and $\widehat{\pi}^{\hat{\mathtt{p}}} : \mathbf{L}^2(\mathscr{E}) \to \widehat{\mathbf{V}}_h$, the L^2 orthogonal projections onto the discrete spaces. The boundary value problem (1)–(2) is then approximated by the following finite element scheme.

Method 1. Find $\mathbf{u}_h \in \mathbf{V}_h$, $\widehat{\mathbf{u}}_h \in \widehat{\mathbf{V}}_h$, and $p_h \in Q_h$, such that

$$\begin{cases} \mathbf{a}_h(\mathbf{u}_h, \widehat{\mathbf{u}}_h; \mathbf{v}_h, \widehat{\mathbf{v}}_h) + \mathbf{b}_h(\mathbf{v}_h, \widehat{\mathbf{v}}_h; p_h) = (\mathbf{f}, \mathbf{v}_h)_{\mathscr{T}_h}, \\ \mathbf{b}_h(\mathbf{u}_h, \widehat{\mathbf{u}}_h; q_h) \qquad\qquad\qquad\quad = (g, q_h)_{\mathscr{T}_h}, \end{cases}$$

for all $\mathbf{v}_h \in \mathbf{V}_h$, $\widehat{\mathbf{v}}_h \in \widehat{\mathbf{V}}_h$, and $q_h \in Q_h$. The bilinear forms are defined as

$$\mathbf{a}_h(\mathbf{u}, \widehat{\mathbf{u}}; \mathbf{v}, \widehat{\mathbf{v}}) := \sigma \mathbf{d}_h(\mathbf{u}, \widehat{\mathbf{u}}; \mathbf{v}, \widehat{\mathbf{v}}) + 2\mu \mathbf{s}_h(\mathbf{u}, \widehat{\mathbf{u}}; \mathbf{v}, \widehat{\mathbf{v}}),$$
$$\mathbf{b}_h(\mathbf{v}, \widehat{\mathbf{v}}; q) := -(\operatorname{div} \mathbf{v}, q)_{\mathscr{T}_h} + \langle \mathbf{v} - \widehat{\mathbf{v}}, q\mathbf{n} \rangle_{\partial \mathscr{T}_h},$$

and the bilinear forms $\mathbf{d}_h$ and $\mathbf{s}_h$ are given by

$$\mathbf{d}_h(\mathbf{u}, \widehat{\mathbf{u}}; \mathbf{v}, \widehat{\mathbf{v}}) := (\mathbf{u}, \mathbf{v})_{\mathscr{T}_h} + \alpha \langle (\widehat{\pi}^{\hat{\mathtt{p}}}\mathbf{u} - \widehat{\mathbf{u}}) \cdot \mathbf{n}, (\widehat{\pi}^{\hat{\mathtt{p}}}\mathbf{v} - \widehat{\mathbf{v}}) \cdot \mathbf{n} \rangle_{\partial \mathscr{T}_h},$$
$$\begin{aligned} \mathbf{s}_h(\mathbf{u}, \widehat{\mathbf{u}}; \mathbf{v}, \widehat{\mathbf{v}}) := {} & (\varepsilon(\mathbf{u}), \varepsilon(\mathbf{v}))_{\mathscr{T}_h} - \langle \varepsilon(\mathbf{u}) \cdot \mathbf{n}, \mathbf{v} - \widehat{\mathbf{v}} \rangle_{\partial \mathscr{T}_h} \\ & - \langle \mathbf{u} - \widehat{\mathbf{u}}, \varepsilon(\mathbf{v}) \cdot \mathbf{n} \rangle_{\partial \mathscr{T}_h} + \alpha \langle \widehat{\pi}^{\hat{\mathtt{p}}}\mathbf{u} - \widehat{\mathbf{u}}, \widehat{\pi}^{\hat{\mathtt{p}}}\mathbf{v} - \widehat{\mathbf{v}} \rangle_{\partial \mathscr{T}_h}. \end{aligned}$$

One easily verifies that any regular solution of (1)–(2) also satisfies the discrete variational principle above.

Proposition 1 (Consistency). *Let* $(\mathbf{u}, p)$ *denote a solution of* (1)–(2)*, and assume additionally that* $\mathbf{u} \in \mathbf{H}^2(\mathscr{T}_h)$ *and* $p \in H^1(\mathscr{T}_h)$*. Then*

$$\mathbf{a}_h(\mathbf{u}, \mathbf{u}; \mathbf{v}_h, \widehat{\mathbf{v}}_h) + \mathbf{b}_h(\mathbf{v}_h, \widehat{\mathbf{v}}_h; p) = (\mathbf{f}, \mathbf{v}_h)_{\mathscr{T}_h} \quad \textit{and} \quad \mathbf{b}_h(\mathbf{u}, \mathbf{u}; q_h) = (g, q_h)_{\mathscr{T}_h}$$

for all $\mathbf{v}_h \in \mathbf{V}_h$, $\widehat{\mathbf{v}}_h \in \widehat{\mathbf{V}}_h$, *and* $q_h \in Q_h$*; thus, Method 1 is consistent.*

In the Darcy limit $\mu = 0$, it suffices to require $\mathbf{u} \in \mathbf{H}^1(\mathscr{T}_h)$.

4 Stability and Error Analysis

An important ingredient for our analysis will be the following result.

Lemma 1 (Discrete Korn inequality). *Let* $\hat{\mathrm{p}} \geq 1$. *Then there is a* $\kappa > 0$ *independent of* h, *such that for all* $\mathbf{v} \in \mathbf{H}^1(\mathscr{T}_h)$ *and* $\widehat{\mathbf{v}} \in \mathbf{L}^2(\mathscr{E})$, *there holds*

$$\|\varepsilon(\mathbf{v})\|^2_{\mathscr{T}_h} + |\widehat{\pi}^{\hat{\mathrm{p}}}(\mathbf{v} - \widehat{\mathbf{v}})|^2_{1/2,\partial\mathscr{T}_h} \geq \kappa \|\nabla \mathbf{v}\|^2_{\mathscr{T}_h}. \tag{5}$$

Proof. The statement follows via the triangle inequality from Korn's inequality for piecewise H^1 functions [3, Eq. (1.12)] established by Brenner. □

Proposition 2. *For any* $(\mathbf{v}_h, \widehat{\mathbf{v}}_h) \in \mathbf{V}_h \times \widehat{\mathbf{V}}_h$ *there holds*

$$\mathbf{s}_h(\mathbf{v}_h, \widehat{\mathbf{v}}_h; \mathbf{v}_h, \widehat{\mathbf{v}}_h) \geq \min\{\tfrac{5}{12}, \tfrac{\kappa}{4}\}\big(\|\nabla \mathbf{u}\|^2_{\mathscr{T}_h} + |\widehat{\pi}^{\hat{\mathrm{p}}}(\mathbf{u} - \widehat{\mathbf{u}})|^2_{1/2,\partial\mathscr{T}_h}\big).$$

Proof. By Young's inequality, Eq. (3) and (4), we obtain

$$-2\langle \varepsilon(\mathbf{v}_h)\cdot\mathbf{n}, \mathbf{v}_h - \widehat{\mathbf{v}}_h\rangle_{\partial T} \geq -\tfrac{3}{4}\|\varepsilon(\mathbf{v}_h)\|^2_T - \tfrac{1}{3}|\widehat{\pi}^{\hat{\mathrm{p}}}(\mathbf{v}_h - \widehat{\mathbf{v}}_h)|^2_{1/2,\partial T}.$$

The result then follows by Lemma 1, and the definition of $\mathbf{s}_h$. □

For appropriately characterizing the coercivity of the bilinear form $\mathbf{d}_h$, let us introduce the discrete kernel space for the bilinear form $\mathbf{b}_h$, namely

$$\mathbf{K}_h := \{(\mathbf{v}_h, \widehat{\mathbf{v}}_h) \in \mathbf{V}_h \times \widehat{\mathbf{V}}_h : \mathbf{b}_h(\mathbf{v}_h, \widehat{\mathbf{v}}_h; q_h) = 0 \ \forall q_h \in Q_h\}.$$

Proposition 3. *For any pair of functions* $(\mathbf{v}_h, \widehat{\mathbf{v}}_h) \in \mathbf{K}_h$ *there holds*

$$\mathbf{d}_h(\mathbf{v}_h, \widehat{\mathbf{v}}_h; \mathbf{v}_h, \widehat{\mathbf{v}}_h) \geq \|\mathbf{v}_h\|^2_{\mathscr{T}_h} + \|\operatorname{div}\mathbf{v}_h\|^2_{\mathscr{T}_h} + \tfrac{3}{4}|\widehat{\pi}^{\hat{\mathrm{p}}}(\mathbf{v}_h - \widehat{\mathbf{v}}_h)\cdot\mathbf{n}|^2_{1/2,\partial\mathscr{T}_h}.$$

Proof. Note that for every $T \in \mathscr{T}_h$ we have $\operatorname{div}\mathbf{v}_h|_T \in \mathscr{P}_{\mathrm{q}}(T)$. Testing with $q_h = \operatorname{div}\mathbf{v}_h$ and using (3) yields

$$\|\operatorname{div}\mathbf{v}_h\|^2_T = \langle(\mathbf{v}_h - \widehat{\mathbf{v}}_h)\cdot\mathbf{n}, \operatorname{div}\mathbf{v}_h\rangle_{\partial T} \leq \tfrac{1}{2}|(\widehat{\pi}^{\hat{\mathrm{p}}}\mathbf{v}_h - \widehat{\mathbf{v}}_h)\cdot\mathbf{n}|_{1/2,\partial T}\|\operatorname{div}\mathbf{v}_h\|_T,$$

and hence $\|\operatorname{div}\mathbf{v}_h\|_{\mathscr{T}_h} \leq \frac{1}{2}|(\widehat{\pi}^{\hat{\mathrm{p}}}\mathbf{v}_h - \widehat{\mathbf{v}}_h)\cdot\mathbf{n}|_{1/2,\partial\mathscr{T}_h}$. The result then follows by adding and subtracting $\|\operatorname{div}\mathbf{v}_h\|^2_{\partial\mathscr{T}_h}$ from the bilinear form $\mathbf{d}_h$. □

The two coercivity estimates suggest to utilize the following energy norms for the stability analysis of Method 1, namely, $\|q\|_{0,\mathscr{T}_h}$ and

$$\begin{aligned}\|(\mathbf{v}, \widehat{\mathbf{v}})\|^2_{1,\mathscr{T}_h} := \sigma\big(\|\mathbf{v}\|^2_{\mathscr{T}_h} + \|\operatorname{div}\mathbf{v}\|^2_{\mathscr{T}_h} + |\widehat{\pi}^{\hat{\mathrm{p}}}(\mathbf{v} - \widehat{\mathbf{v}})\cdot\mathbf{n}|^2_{1/2,\partial\mathscr{T}_h}\big) \\ + \mu\big(\|\nabla\mathbf{v}\|^2_{\mathscr{T}_h} + |\widehat{\pi}^{\hat{\mathrm{p}}}(\mathbf{v} - \widehat{\mathbf{v}})|^2_{1/2,\partial\mathscr{T}_h}\big).\end{aligned}$$

Remark 1. If $\mu = 0$, then $\|(\cdot,\cdot)\|_{1,\mathscr{T}_h}$ is only a semi-norm on $\mathbf{V}_h \times \widehat{\mathbf{V}}_h$. This deficiency can be overcome by eliminating the tangential velocities in the definition of the hybrid space, or by penalizing also the jump of the tangential velocities in the bilinear form $\mathbf{d}_h$. Both remedies do not affect our analysis.

A combination of Propositions 2 and 3 now yields the kernel ellipticity for $\mathbf{a}_h$.

Proposition 4 (Coercivity). *For any element* $(\mathbf{v}_h, \widehat{\mathbf{v}}_h) \in \mathbf{K}_h$ *there holds*

$$\mathbf{a}_h(\mathbf{v}_h, \widehat{\mathbf{v}}_h; \mathbf{v}_h, \widehat{\mathbf{v}}_h) \geq \min\{\tfrac{3}{4}, \tfrac{\kappa}{2}\} \|(\mathbf{v}_h, \widehat{\mathbf{v}}_h)\|^2_{1,\mathscr{T}_h}.$$

The constants C_i appearing in the following results depend on the bounds m and M, but are else independent of the parameters μ, σ, and of $\mathtt{h}$ and $\mathtt{p}$. Let us next consider the operator $(\pi^{\mathtt{p}}, \widehat{\pi}^{\hat{\mathtt{p}}}) : \mathbf{H}^1_0(\Omega) \to \mathbf{V}_h \times \widehat{\mathbf{V}}_h$.

Proposition 5 (Fortin operator). *For any* $\mathbf{v} \in \mathbf{H}^1_0(\Omega)$ *there holds*

$$b_h(\pi^{\mathtt{p}}\mathbf{v}, \widehat{\pi}^{\hat{\mathtt{p}}}\mathbf{v}; q_h) = b(\mathbf{v}, q_h) \quad \forall q_h \in Q_h, \tag{6}$$

and
$$\|(\pi^{\mathtt{p}}\mathbf{v}, \widehat{\pi}^{\hat{\mathtt{p}}}\mathbf{v})\|_{1,\mathscr{T}_h} \leq C_\pi \mathtt{p}^{1/2} \|\mathbf{v}\|_{1,\Omega}. \tag{7}$$

Proof. Equation (6) follows from the properties of the projections, and (7) results from stability estimates for the L^2 projections; cf. [7] for details. □

The inf-sup stability of $\mathbf{b}_h$ now follows directly from the previous result.

Proposition 6 (Inf-sup condition). *For any* $q_h \in Q_h$ *there holds*

$$\sup_{(\mathbf{v}_h, \widehat{\mathbf{v}}_h) \in \mathbf{V}_h \times \widehat{\mathbf{V}}_h} \frac{\mathbf{b}_h(\mathbf{v}_h, \widehat{\mathbf{v}}_h; q_h)}{\|(\mathbf{v}_h, \widehat{\mathbf{v}}_h)\|_{1,\mathscr{T}_h}} \geq C_\beta \mathtt{p}^{-1/2} \|q_h\|_{0,\mathscr{T}_h}. \tag{8}$$

As a consequence of Propositions 4 and 6, we obtain by Brezzi's theorem that Method 1 has a unique solution and thus is well-defined. Next, we show the boundedness of the bilinear forms with respect to a pair of stronger norms defined by $|||q_h|||^2_{0,\mathscr{T}_h} := \|q_h\|^2_{\mathscr{T}_h} + |q_h \cdot \mathbf{n}|^2_{-1/2,\partial\mathscr{T}_h}$ and

$$|||(\mathbf{v}_h, \widehat{\mathbf{v}}_h)|||^2_{1,\mathscr{T}_h} := \|(\mathbf{v}_h, \widehat{\mathbf{v}}_h)\|^2_{1,\mathscr{T}_h} + \mu |\partial_{\mathbf{n}} \mathbf{v}_h|^2_{-1/2,\partial\mathscr{T}_h},$$

The norms $\|\cdot\|_{0,\mathscr{T}_h}$, $\|(\cdot,\cdot)\|_{1,\mathscr{T}_h}$ and $|||\cdot|||_{0,\mathscr{T}_h}$, $|||(\cdot,\cdot)|||_{1,\mathscr{T}_h}$ are equivalent on the finite element spaces with equivalence constants less than two. This yields coercivity and inf-sup stability of $\mathbf{a}_h$ and $\mathbf{b}_h$ also with respect to the stronger norms. The following bounds then follow from the Cauchy-Schwarz inequality.

Proposition 7 (Boundedness). *For any* $\widehat{\mathbf{u}}$, $\widehat{\mathbf{v}} \in \widehat{\mathbf{V}}_h \oplus \mathbf{L}^2(\mathscr{E})$ *and every function* $\mathbf{u}$, $\mathbf{v} \in \mathbf{V}_h \oplus (\mathbf{H}^1_0(\Omega) \cap \mathbf{H}^2(\mathscr{T}_h))$, *there holds*

$$\mathbf{a}_h(\mathbf{u}, \widehat{\mathbf{u}}; \mathbf{v}, \widehat{\mathbf{v}}) \leq C_a |||(\mathbf{u}, \widehat{\mathbf{u}})|||_{1,\mathscr{T}_h} |||(\mathbf{v}, \widehat{\mathbf{v}})|||_{1,\mathscr{T}_h},$$

and for all $p \in Q_h \oplus (L^2_0(\Omega) \cap H^1(\mathscr{T}_h))$, *there holds additionally*

$$\mathbf{b}_h(\mathbf{u}, \widehat{\mathbf{u}}; p) \leq C_b |||(\mathbf{u}, \widehat{\mathbf{u}})|||_{1,\mathscr{T}_h} |||p|||_{0,\mathscr{T}_h}.$$

Standard polynomial approximation results [2] imply the following properties.

Proposition 8 (Approximation). *Assume $s \geq 1$. Then for any function $\mathbf{u} \in \mathbf{H}^1_0(\Omega) \cap \mathbf{H}^{s+1}(\mathscr{T}_h)$ there exist elements $\mathbf{v}_h \in \mathbf{V}_h$ and $\widehat{\mathbf{v}}_h \in \widehat{\mathbf{V}}_h$ such that*

$$|||(\mathbf{u}-\mathbf{v}_h, \mathbf{u}-\widehat{\mathbf{v}}_h)|||_{1,\mathscr{T}_h} \leq C_{ap}\, \mathtt{p}^{1/2-s} h^{\min\{\mathtt{p},s\}} \|\mathbf{u}\|_{s+1,\mathscr{T}_h},$$

and for any $p \in L^2_0(\Omega) \cap H^s(\mathscr{T}_h)$ there exists a $q_h \in Q_h$ such that

$$|||p - q_h|||_{0,\mathscr{T}_h} \leq C_{ap}\, \mathtt{p}^{-s} h^{\min\{s,\mathtt{q}+1\}} \|p\|_{s,\mathscr{T}_h}.$$

The a-priori estimates now follow by combination of the previous results.

Proposition 9 (Error estimate). *Let $(\mathbf{u}, p)$ be the solution of (1)–(2), and let $(\mathbf{u}_h, \widehat{\mathbf{u}}_h, p_h)$ denote the approximation defined by Method 1. Then*

$$\begin{aligned} |||(\mathbf{u}-\mathbf{u}_h, \mathbf{u}-\widehat{\mathbf{u}}_h)|||_{1,\mathscr{T}_h} + \mathtt{p}^{-1/2} |||p - p_h|||_{0,\mathscr{T}_h} \\ \leq C_{err}\, \mathtt{p}^{1/2} h^{\min\{\mathtt{p},s\}} \big(\mathtt{p}^{1/2-s} \|\mathbf{u}\|_{s+1,\mathscr{T}_h} + \mathtt{p}^{-s} \|p\|_{s,\mathscr{T}_h} \big). \end{aligned}$$

Proof. The result follows with the usual arguments from the consistency, discrete stability, and boundedness of the bilinear forms, and the approximation properties of the finite element spaces; for details, see [7] or [9].

5 Remarks

The analysis of Sect. 4 applies almost verbatim to spatially varying material parameters μ and σ. In particular, a coupling of Darcy and Stokes equations in different parts of the domain is possible and treated automatically. A numerical example for such a case is presented in the next section.

Our results can be extended to shape regular meshes and varying polynomial degree. Also meshes with a bounded number of hanging nodes on each edge or face, and even more general non-conforming mortar meshes can be treated. We refer to [6, 7] for a detailed discussion of conditions on the mesh and polynomial degree distribution.

The coercivity and boundedness estimates also hold for more general finite element spaces, but we explicitly utilized the complete discontinuity of the spaces in the proof of the inf-sup condition. Other constructions of a Fortin-operator, cf. e.g. [9], would allow to relax this assumption.

Our analysis also covers equal order approximations $\mathtt{q} = \mathtt{p}$, which are stabilized sufficiently by the jump penalty terms.

All degrees of freedom except the piecewise constant pressure and the hybrid velocities can be eliminated by static condensation on the element level. This leads to small global systems, which for $\hat{\mathtt{p}} = 0$ exhibit the same sparsity pattern as a non-conforming $P_1 - P_0$ discretization. For $\hat{\mathtt{p}} = 0$, the discrete Korn inequality (5) is not valid, so this choice had to be excluded here. If $\varepsilon(\mathbf{u})$ in (1) is replaced by $\frac{1}{2}\nabla\mathbf{u}$, we however obtain a stable scheme.

6 Numerical Results

Let us now illustrate the capability of the proposed method to deal with incompressible flow in various regimes. Our numerical results were obtained with an implementation of Method 1 in NGSolve.[3]

As a first example, we consider the generalized Stokes equation (1) on the unit square $\Omega = (-1,1)^2$ with a known analytic solution given by

$$\mathbf{u} = \left(20xy^3, 5x^4 - 5y^4\right), \quad p = 60x^2y - 20y^3, .$$

The data $\mathbf{f}$ and g can be obtained from Eq. (1). For the numerical solution, we employed Method 1 with $\mathtt{p} = \hat{\mathtt{p}} = 2$ and $\mathtt{q} = 1$ on a sequence of uniformly refined meshes for different values of μ and σ. The analytic solution was used to compute the errors listed in Table 1. As predicted by the theory, we can observe the optimal quadratic convergence.

Table 1. Energy errors obtained by simulation on a sequence of uniformly refined meshes for $(\sigma,\mu) \in \{(1,0),(\frac{1}{2},\frac{1}{2}),(0,1)\}$, resembling Darcy, Brinkman, and Stokes flow.

level	Darcy	rate	Brinkman	rate	Stokes	rate
0	4.3996	–	3.4058	–	3.8578	–
1	1.1261	1.96	0.8628	1.98	0.9764	1.98
2	0.2799	2.00	0.2146	2.00	0.2428	2.00
3	0.0678	2.04	0.0533	2.00	0.0603	2.00

As a second test case, we consider a coupled Darcy-Stokes flow on a domain consisting of two subdomains Ω_D and Ω_S, as depicted in Fig. 1. The flow in the subdomains is governed by

$$\sigma_i \mathbf{u}_i - 2\mu_i \operatorname{div} \varepsilon(\mathbf{u}_i) + \nabla p_i = 0 \quad \text{and} \quad \operatorname{div} \mathbf{u}_i = 0 \qquad \text{in } \Omega_i,$$

with $\mu_D = 0$ in the Darcy domain Ω_D, and $\sigma_S = 0$ in the Stokes domain Ω_S, and the subproblems are coupled across the interface $\partial\Omega_D \cap \partial\Omega_S$ through

$$\mathbf{u}_S \cdot \mathbf{n} = \mathbf{u}_D \cdot \mathbf{n}, \;\; p_S - 2\mu(\varepsilon(\mathbf{u}_S)\cdot\mathbf{n})\cdot\mathbf{n} = p_D, \;\; \mathbf{u}_S \cdot \tau + 2\gamma(\varepsilon(\mathbf{u}_S)\cdot\mathbf{n})\cdot\tau = 0.$$

For $\gamma = 0$, these conditions arise naturally when considering the generalized Stokes problem (1) with discontinuous coefficients. In the case $\gamma \neq 0$ the third *Beaver-Joseph-Saffman* condition, which restricts the tangential components of the normal stresses, gives rise to an additional interface term that has to be included in the definition of the bilinear form $\mathbf{a}_h$; for details see [8] and the references given there.

Acknowledgments This work was supported by the German Research Association (DFG) through grant GSC 111.

[3] visit: http://sourceforge.net/apps/mediawiki/ngsolve

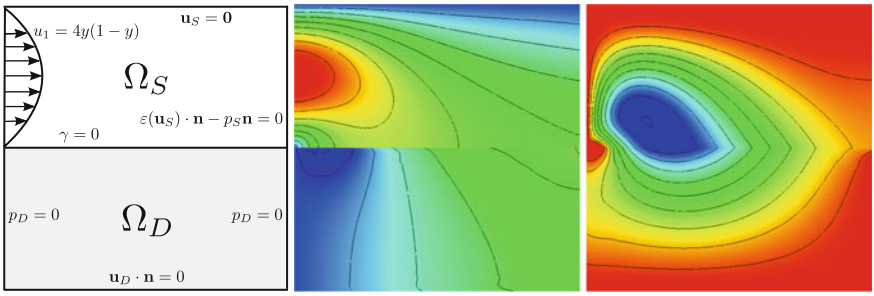


Fig. 1. From *left* to *right*: problem setup, and isolines of x- and y-components of the velocity for parameters $\mu_S = 1$, $\sigma_S = 0$ and $\mu_D = 0$, σ_D=1; $\gamma = 0$. A part of the flow soaks through the porous medium. The normal component of the velocity is (almost) continuous across the interface, while no continuity is obtained for the tangential component

Bibliography

[1] D. N. Arnold, F. Brezzi, B. Cockburn, and D. Marini. Unified analysis of discontinuous Galerkin methods for elliptic problems. *SIAM J. Numer. Anal.*, 39:1749–1779, 2002.

[2] I. Babuška and M. Suri. The *hp* version of the finite element method with quasiuniform meshes. *M2AN*, 21:199–238, 1987.

[3] S. C. Brenner. Korn's inequalities for piecewise H^1 vector fields. *Math. Comp.*, 73(247):1067–1087, 2004.

[4] E. Burman and P. Hansbo. A unified stabilized method for Stokes' and Darcy's equations. *J. Comput. Appl. Math.*, 198:35–51, 2007.

[5] B. Cockburn, J. Gopalakrishnan, and R. Lazarov. Unified hybridization of discontinuous Galerkin, mixed and continuous Galerkin methods for second order elliptic problems. *SIAM J. Numer. Anal.*, 47:1319–1365, 2009.

[6] H. Egger and C. Waluga. *hp*-Analysis of a Hybrid DG Method for Stokes Flow. *IMA Journal of Numerical Analysis* 2012; doi: 10.1093/imanum/drs018.

[7] H. Egger and C. Waluga. A Hybrid Mortar Method for Incompressible Flow. *International Journal of Numerical Analysis and Modeling*, 9(4):793–812, 2012.

[8] G. Kanschat and B. Riviére. A strongly conservative finite element method for the coupling of Stokes and Darcy flow. *J. Comput. Phys.*, 229:5933–5943, 2010.

[9] D. Schötzau, C. Schwab, and A. Toselli. Mixed hp-DGFEM for incompressible flows. *SIAM J. Numer. Anal.*, pages 2171–2194, 2003.

A Parallel Monolithic Domain Decomposition Method for Blood Flow Simulations in 3D

Yuqi Wu[1] and Xiao-Chuan Cai[2]

[1] Department of Applied Mathematics, University of Colorado at Boulder, Boulder, CO 80309, USA, yuqi.wu@colorado.edu
[2] Department of Computer Science, University of Colorado at Boulder, Boulder, CO 80309, USA, cai@cs.colorado.edu

Summary. We develop a parallel scalable domain decomposition method for the simulation of blood flows in compliant arteries in 3D, by using a fully coupled system of linear elasticity equation and incompressible Navier-Stokes equations. The system is discretized with a finite element method on unstructured moving meshes and solved by a Newton-Krylov algorithm preconditioned with an overlapping additive Schwarz method. We focus on the accuracy and parallel scalability of the algorithm, and report the parallel performance and robustness of the proposed approach by some numerical experiments carried out on a supercomputer with a large number of processors and for problems with millions of unknowns.

1 Introduction

Computer modeling of fluid-structure interaction (FSI) is a useful tool for the study of hemodynamics of blood flows in human arteries. Accurate modeling helps the prediction and treatment of, for example, vascular diseases. FSI problems are in general difficult to study. One of the main challenges is the effective coupling of the fluid and the structure. Two well-known formulations are iterative and monolithic. In iterative approaches, the fluid and the structure equations are solved one after the other repeatedly, until some desired tolerance is reached [7, 10]. The convergence of these approaches is difficult to achieve in some situations [6], since the approaches are very similar to nonlinear Gauss-Seidel with two large blocks. In contrast, we develop a monolithic coupling similar to [2–4], where the fluid and the structure equations are solved simultaneously in a fully coupled fashion and the coupling conditions are enforced strongly as part of the system. The monolithic approach has been shown to be more robust. Many of the convergence problems encountered within the iterative approaches can be avoided.

With the rapid advancement in high performance computing technologies, high resolution blood flow simulations are expected to provide more details of the physics of blood flows and the artery walls. To obtain highly accurate solutions on a very fine mesh, the parallel performance and scalability of the solution algorithm is becoming

R. Bank et al. (eds.), *Domain Decomposition Methods in Science and Engineering XX*, Lecture Notes in Computational Science and Engineering 91,
DOI 10.1007/978-3-642-35275-1_80,

a key issue in the simulation. In [2, 3], a class of parallel scalable Newton-Krylov-Schwarz method was introduced for FSI in 2D. In this paper, we focus on solving the fully coupled FSI system in 3D and also discuss the parallel performance and robustness of the algorithms. The rest of the paper is organized as follows. In Sect. 2, we describe the formulation and the discretization of the fully coupled FSI problem. In Sect. 3, we present the Newton-Krylov-Schwarz method for solving the fully coupled nonlinear system. In Sect. 4, we first validate the method by comparing solutions obtained with the new approach with published results for a straight cylinder problem, then report the parallel performance of the algorithm. Finally, we provide some concluding remarks in Sect. 5.

2 Mathematical Formulation and Discretization

Our fully coupled approach can be described by the coupling of three components, the linear elasticity equation for the wall structure in the reference Lagrangian frame, the incompressible Navier-Stokes equations for the fluid in the arbitrary Lagrangian-Eulerian (ALE) framework, and the Laplace equation for the displacement of the fluid domain.

Let $\Omega_s \in R^3$ be the structure domain. The displacement $\mathbf{x}_s$ of the artery walls is described by

$$\rho_s \frac{\partial^2 \mathbf{x}_s}{\partial t^2} - \nabla \cdot \sigma_s = \mathbf{f}_s \quad \text{in } \Omega_s, \tag{1}$$

where ρ_s is the density of the structure, and $\sigma_s = \lambda_s(\nabla \cdot \mathbf{x}_s)I + \mu_s(\nabla \mathbf{x}_s + \nabla \mathbf{x_s}^T)$ is the Cauchy stress tensor. The Lamé parameters λ_s and μ_s are related to the Young's modulus E and the Poisson ratio ν_s by $\lambda_s = \nu_s E/((1+\nu_s)(1-2\nu_s))$ and $\mu_s = E/(2(1+\nu_s))$. We fix the structure displacement $\mathbf{x}_s = 0$ on the inlet and outlet boundary Γ_s, and apply the zero normal traction condition $\sigma_s \cdot \mathbf{n} = 0$ on the external boundaries.

In order to model the fluid in a moving domain $\Omega_f(t) \in R^3$, the displacement of the fluid domain $\mathbf{x}_f$ in the reference configuration $\Omega_0 \in R^3$ is assumed to satisfy a Laplace equation,

$$\Delta \mathbf{x}_f = 0 \quad \text{in } \Omega_0.$$

We define an ALE mapping A_t from Ω_0 to $\Omega_f(t)$:

$$A_t : \Omega_0 \to \Omega_f(t), \quad A_t(\mathbf{Y}) = \mathbf{Y} + \mathbf{x_f}(\mathbf{Y}), \quad \forall \mathbf{Y} \in \Omega_0,$$

where $\mathbf{Y}$ is referred to as the ALE coordinates. The incompressible Navier-Stokes equations defined on the moving domain $\Omega_f(t)$ are written in the ALE form as

$$\rho_f \left.\frac{\partial \mathbf{u}_f}{\partial t}\right|_{\mathbf{Y}} + \rho_f[(\mathbf{u}_f - \omega_g) \cdot \nabla]\mathbf{u}_f = \nabla \cdot \sigma_f \qquad \text{in } \Omega_f(t),$$

$$\nabla \cdot \mathbf{u}_f = 0 \qquad \text{in } \Omega_f(t),$$

where ρ_f is the fluid density, $\mathbf{u}_f$ is the fluid velocity, and $\sigma_f = -p_f I + \mu_f(\nabla \mathbf{u_f} + \nabla \mathbf{u_f}^T)$ is the Cauchy stress tensor. $\omega_g = \partial \mathbf{x_f}/\partial t$ is the velocity of the moving domain and $\mathbf{Y}$ indicates that the time derivative is taken with respect to the ALE coordinates. On the inlet boundary Γ_i, a given velocity profile is prescribed. On the outlet boundary Γ_o, the zero traction condition $\sigma_f \cdot \mathbf{n} = 0$ is considered, where $\mathbf{n}$ is the unit outward normal. These boundary conditions may be chosen differently, depending on the problem at hand.

More importantly, three coupling conditions are strongly enforced on the fluid-structure interface Γ_w

$$\sigma_s \cdot \mathbf{n}_s = -\sigma_f \cdot \mathbf{n}_f, \;\; \mathbf{u_f} = \frac{\partial \mathbf{x_s}}{\partial t}, \;\; \mathbf{x}_f = \mathbf{x}_s, \tag{2}$$

where $\mathbf{n}_s$, $\mathbf{n}_f$ are unit normal vectors on the fluid-structure interface.

By introducing the structure velocity $\dot{\mathbf{x}}_\mathbf{s}$ as an additional unknown variable, we can rewrite the structure momentum equation (1) as a first-order system of equations. We define the variational space of the structure problem as

$$X = \left\{ \mathbf{x_s} \in [H^1(\Omega_s)]^3 : \mathbf{x_s} = 0 \text{ on } \Gamma_s \right\}.$$

The weak form of the structure problem is stated as follows: Find $\mathbf{x_s} \in X$ and $\dot{\mathbf{x}}_\mathbf{s} \in X$ such that $\forall \phi_s \in X$ and $\forall \varphi_s \in X$,

$$\begin{aligned} B_s(\{\mathbf{x_s}, \dot{\mathbf{x}}_\mathbf{s}\}, \{\phi_s, \varphi_s\}; \sigma_f) &= \rho_s \frac{\partial}{\partial t} \int_{\Omega_s} \dot{\mathbf{x}}_\mathbf{s} \cdot \phi_s \, d\Omega + \int_{\Omega_s} \nabla \phi_s : \sigma_s \, d\Omega \\ - \int_{\Gamma_w} \phi_s \cdot \left(\sigma_f \cdot \mathbf{n}_s\right) \, ds - \int_{\Omega_s} \mathbf{f}_s \cdot \phi_s \, d\Omega &+ \int_{\Omega_s} \left(\frac{\partial \mathbf{x_s}}{\partial t} - \dot{\mathbf{x}}_\mathbf{s} \right) \cdot \varphi_s \, d\Omega = 0. \end{aligned}$$

The variational spaces of the fluid subproblem are time dependent, and the solution of the structure subproblem provides an essential boundary condition for the fluid subproblem by (2). We define the trial and weighting function spaces as:

$$\begin{aligned} V &= \left\{ \mathbf{u}_f \in [H^1(\Omega_f(t))]^3 : \mathbf{u}_f = g \text{ on } \Gamma_i, \mathbf{u}_f = \partial \mathbf{x}_s / \partial t \text{ on } \Gamma_w \right\}, \\ V_0 &= \left\{ \mathbf{u}_f \in [H^1(\Omega_f(t))]^3 : \mathbf{u}_f = 0 \text{ on } \Gamma_i \cup \Gamma_w \right\}, \\ P &= L^2\left(\Omega_f(t)\right). \end{aligned}$$

The weak form of the fluid problem reads: Find $\mathbf{u}_f \in V$ and $p_f \in P$ such that $\forall \phi_f \in V_0$ and $\forall \psi_f \in P$,

$$\begin{aligned} B_f(\{\mathbf{u}_f, p_f\}, \{\phi_f, \psi_f\}; \mathbf{x}_f) &= \rho_f \int_{\Omega_f(t)} \left. \frac{\partial \mathbf{u}_f}{\partial t} \right|_\mathbf{Y} \cdot \phi_f \, d\Omega - \int_{\Omega_f(t)} p_f (\nabla \cdot \phi_f) \, d\Omega \\ &+ \rho_f \int_{\Omega_f(t)} \left[(\mathbf{u}_f - \omega_g) \cdot \nabla \right] \mathbf{u}_f \cdot \phi_f \, d\Omega + 2\mu_f \int_{\Omega_f(t)} \varepsilon(\mathbf{u}_f) : \varepsilon(\phi_f) \, d\Omega \\ &+ \int_{\Omega_f(t)} (\nabla \cdot \mathbf{u}_f) \psi_f \, d\Omega = 0, \end{aligned}$$

where $\varepsilon(\mathbf{u}_f) = (\nabla \mathbf{u}_f + \nabla \mathbf{u}_f^T)/2$.

The weak form of the domain movement problem reads: Find $\mathbf{x}_f \in Z$ such that $\forall \xi \in Z_0$,

$$B_m(\mathbf{x}_f, \xi) = \int_{\Omega_0} \nabla \xi : \nabla \mathbf{x}_f \, d\Omega = 0.$$

And the variational spaces are defined as

$$Z_0 = \{\mathbf{x}_f \in [H^1(\Omega_0)]^3 : \mathbf{x}_f = 0 \text{ on } \Gamma_i \cup \Gamma_o \cup \Gamma_w\},$$
$$Z = \{\mathbf{x}_f \in [H^1(\Omega_0)]^3 : \mathbf{x}_f = \mathbf{x}_s \text{ on } \Gamma_w, \mathbf{x}_f = 0 \text{ on } \Gamma_i \cup \Gamma_o\}.$$

We discretize the fully coupled problem in space with a finite element method, by using unstructured P1-P1 stabilized elements for the fluid, P1 elements for the structure and P1 elements for the fluid domain motion. We denote the finite element subspaces X_h, V_h, $V_{h,0}$, P_h, Z_h, $Z_{h,0}$ as the counterparts of their infinite dimensional subspaces. Because the fluid problem requires that the pair V_h and P_h satisfy the LBB inf-sup condition, additional SUPG stabilization terms are needed in the formulation with equal-order interpolation of the velocity and the pressure as described in [11, 12]. The semi-discrete stabilized finite element formulation for the fluid problem reads as follows: Find $\mathbf{u}_f \in V_h$ and $p_f \in P_h$, such that $\forall \phi_f \in V_{h,0}$ and $\forall \psi_f \in P_h$,

$$B\left(\{\mathbf{u}_f, p_f\}, \{\phi_f, \psi_f\}; \mathbf{x}_f\right) = 0,$$

with

$$\begin{aligned} &B\left(\{\mathbf{u}_f, p_f\}, \{\phi_f, \psi_f\}; \mathbf{x}_f\right) \\ &= B_f\left(\{\mathbf{u}_f, p_f\}, \{\phi_f, \psi_f\}; \mathbf{x}_f\right) + \sum_{K \in \mathscr{T}_f^h} \left(\nabla \cdot \mathbf{u}_f, \tau_c \nabla \cdot \phi_f\right)_K \\ &+ \sum_{K \in \mathscr{T}_f^h} \left(\left. \frac{\partial \mathbf{u}_f}{\partial t} \right|_{\mathbf{Y}} + (\mathbf{u}_f - \omega_g) \cdot \nabla \mathbf{u}_f + \nabla p_f, \tau_m \left((\mathbf{u}_f - \omega_g) \cdot \nabla \phi_f + \nabla \psi_f\right) \right)_K, \end{aligned}$$

where $\mathscr{T}_f^h = \{K\}$ is the given unstructured tetrahedral fluid mesh, and τ_c and τ_m are stabilization parameters.

We form the finite dimensional fully coupled FSI problem as follows: Find $x_s \in X_h$, $\dot{x}_s \in X_h$, $u_f \in V_h$, $p_f \in P_h$ and $x_f \in Z_h$ such that $\forall \phi_s \in X_h$, $\forall \varphi_s \in X_h$, $\forall \phi_f \in V_{h,0}$, $\forall \psi_f \in P_h$, and $\forall \xi \in Z_{h,0}$,

$$B_s(\{x_s, \dot{x}_s\}, \{\phi_s, \varphi_s\}; \sigma_f) + B(\{u_f, p_f\}, \{\phi_f, \psi_f\}; x_f) + B_m(x_f, \xi) = 0. \tag{3}$$

The system (3) is further discretized in time with a second-order BDF2 scheme. Since the temporal discretization scheme is fully implicit, at each time step, we obtain the solution x^n at the nth time step from the previous two time steps by solving a sparse, nonlinear algebraic system

$$\mathscr{F}_n(x^n) = 0, \tag{4}$$

where x^n corresponds to the nodal values of the fluid velocity $\mathbf{u}_f$, the fluid pressure p_f, the fluid mesh displacement $\mathbf{x}_f$, the structure displacement $\mathbf{x}_s$ and the structure velocity $\dot{\mathbf{x}}_s$ at the nth time step. For simplicity, we ignore the script n for the rest of the paper.

3 Newton-Krylov-Schwarz Method

In the Newton-Krylov-Schwarz approach, the nonlinear system (4) is solved via the inexact Newton method [8]. At each Newton step the new solution $x^{(k+1)}$ is obtained from the current solution $x^{(k)}$ by $x^{(k+1)} = x^{(k)} + \theta^{(k)} s^{(k)}$, where the step length $\theta^{(k)}$ is determined by a cubic line search technique. The Newton correction $s^{(k)}$ is approximated by solving a preconditioned Jacobian system $J_k M_k^{-1} M_k s^{(k)} = -\mathscr{F}(x^{(k)})$ with GMRES, where M_k^{-1} is a one-level restricted additive Schwarz preconditioner [5].

To define the domain decomposition preconditioner, we first partition the finite element mesh (which consists of the meshes for all components of the coupled system) into non-overlapping subdomains Ω_ℓ^h, $\ell = 1, \ldots, N$, where the number of subdomain N is always the same as the number of processors np. Then, each subdomain Ω_ℓ^h is extended to an overlapping subdomain $\Omega_\ell^{h,\delta}$. Note that the decomposition of the mesh is completely independent of which physical variables are defined on a given mesh point. The number of variables at a given mesh point is considered for the purpose of load balancing. The so-called one-level restricted additive Schwarz preconditioner is defined by

$$M_k^{-1} = \sum_{\ell=1}^{N} (R_\ell^0)^T J_\ell^{-1} R_\ell,$$

where R_ℓ^0 and R_ℓ are restrictions to the degrees of freedom in the non-overlapping subdomain Ω_ℓ^h and the overlapping subdomain $\Omega_\ell^{h,\delta}$, respectively. J_ℓ is a restriction of the Jacobian matrix defined by $J_\ell = R_\ell J_k R_\ell^T$.

4 Numerical Results

Our algorithm is implemented using PETSc [1]. All computations are performed on an IBM BlueGene/L supercomputer.

A benchmark 3D FSI problem is used to study the efficiency and performance of our fully-coupled algorithm and software. The geometry consists of a straight cylinder representing the fluid domain with length 5 cm and radius 0.5 cm, and the surrounding structure with thickness 0.1 cm. A constant traction $\sigma_f \cdot \mathbf{n} = 1.33 \cdot 10^4$ dynes/cm^2 is imposed on the inlet boundary for 3 ms. A zero traction condition is applied to the fluid at the outlet boundary. The fluid is characterized with viscosity $\mu_f = 0.03$ poise, and density $\rho_f = 1.0$ g/cm^3. The Young's modulus $E = 3 \cdot 10^6$ g/(cm s^2), the Poisson ratio $\nu_s = 0.3$, and the structure density $\rho_s = 1.2$ g/cm^3 are the parameters of the structure model.

The fluid and the structure are initially at rest and the simulation is run on a mesh with $2.41 \cdot 10^6$ elements and $3.08 \cdot 10^6$ degrees of freedom, for a total time of 10 ms with a time step size $\Delta t = 0.1$ ms. The simulation proceeds to the next time step when the residual of the nonlinear system is less than 10^{-6}. In Fig. 1, we show the computed fluid pressure and the structure deformation at $t = 2.5, 5.0, 10.0$ ms. Our results are similar to the published results in [7, 9]. We observe that the pressure wave propagates along the cylinder and reaches the end of the cylinder at $t = 10.0$ ms. The wall structure deforms in response to the propagation of the wall pressure, which is a key feature of the fluid-structure interaction.

The strong scalability of the algorithm is presented in Table 1. The results show superlinear scalability for a range of problem sizes and with up to 2,048 processors. It is worth noting that the growth in GMRES iterations for large processor counts may be a problem if we consider to solve the problem on a much larger mesh and with a larger number of processors. In those situations, one possible solution to improve the scalability is the use of a multilevel preconditioner.

Our algorithm is quite robust with respect to physical parameters. In some FSI methods, the convergence becomes difficult to achieve if the density of the fluid and the structure are close to each other. According to Table 2, our solver performs quite well for a wide range of fluid density and structure density.

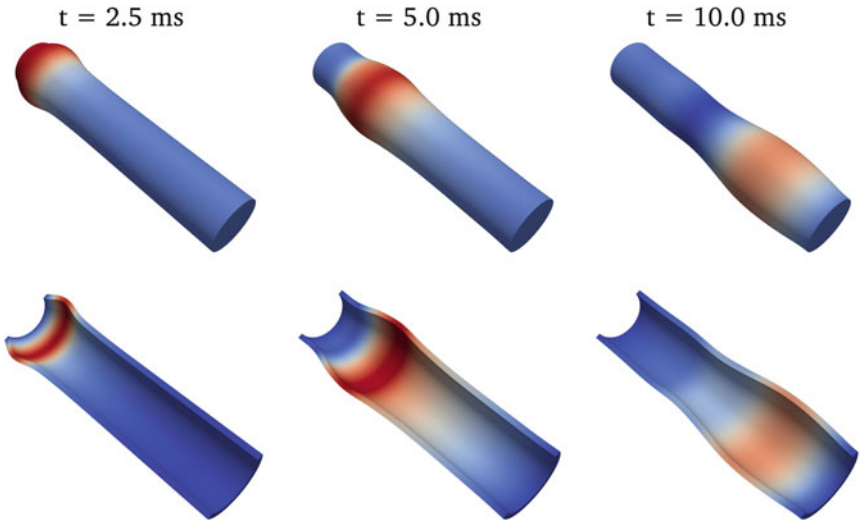


Fig. 1. Pressure wave propagation and structure deformation. The deformation is amplified by a factor of 12 for visualization purpose only

DOF	np	Newton	GMRES	time (s)
$1.24 \cdot 10^6$	256	2.0	41.60	218.03
	512	2.0	49.85	87.53
	1024	2.0	55.65	37.88
$3.07 \cdot 10^6$	512	2.0	57.60	442.44
	1024	2.0	67.15	152.16
	2048	2.0	77.55	65.64

Table 1. Performance with respect to the number of processors for two different mesh sizes. "np" denotes the number of processors. "Newton" denotes the average Newton iteration per time step. "GMRES" denotes the average GMRES iterations per Newton step. "time" refers to the average compute time, in seconds, per time step.

ρ_f	ρ_s	Newton	GMRES	time (s)
1.0	0.1	2.0	71.65	89.94
1.0	1.0	2.0	49.85	87.53
1.0	10.0	2.0	53.90	88.07
1.0	100.0	2.0	61.75	88.84
0.01	1.0	2.0	124.60	96.75
0.1	1.0	2.0	60.90	88.77
10.0	1.0	2.0	60.85	88.79

Table 2. Different combinations of fluid density ρ_f and structure density ρ_s. μ_f is kept at 0.03 poise. The tests are run for a problem with $1.25 \cdot 10^6$ unknowns and 512 processors.

5 Conclusion

In this paper, we developed and studied a parallel scalable overlapping Schwarz domain decomposition method for solving the fully coupled fluid-structure interaction system in 3D. Our algorithm is shown to be scalable on a large scale supercomputer and robust with respect to several important physical parameters.

Bibliography

[1] S. Balay, K. Buschelman, V. Eijkhout, W. Gropp, D. Kaushik, M. Knepley, L. Curfman McInnes, B. Smith, and H. Zhang. PETSc User Manual. Technical report, Argonne National Laboratory, 2010.

[2] A. T. Barker and X.-C. Cai. Scalable parallel methods for monolithic coupling in fluid-structure interaction with application to blood flow modeling. *J. Comput. Phys.*, 229:642–659, 2010.

[3] A. T. Barker and X.-C. Cai. Two-level Newton and hybrid Schwarz preconditioners for fluid-structure interation. *SIAM J. Sci. Comput.*, 32:2395–2417, 2010.

[4] Y. Bazilevs, V. M. Calo, T. J. R. Hughes, and Y. Zhang. Isogeometric fluid-structure interaction: Theory, algorithms and computations. *Comput. Mech.*, 43:3–37, 2008.

[5] X.-C. Cai and M. Sarkis. A restricted additive Schwarz preconditioner for general sparse linear systems. *SIAM J. Sci. Comput.*, 21:792–797, 1999.

[6] P. Causin, J. F. Gerbeau, and F. Nobile. Added-mass effect in the design of partitioned algorithms for fluid-structure problems. *Comput. Methods Appl. Mech. Engrg.*, 194:4506–4627, 2005.

[7] S. Deparis, M. Discacciati, G. Fourestey, and A. Quarteroni. Fluid-structure algorithms based on Steklov-Poincaré operators. *Comput. Methods Appl. Mech. Engrg.*, 195:5797–5812, 2006.

[8] S. C. Eisenstat and H. F. Walker. Choosing the forcing terms in an inexact Newton method. *SIAM J. Sci. Comput.*, 17:16–32, 1996.

[9] M. A. Fernandez and M. Moubachir. A Newton method using exact Jacobians for solving fluid-structure coupling. *Comput. Struct.*, 83:127–142, 2005.

[10] L. Formaggia, J. F. Gerbeau, F. Nobile, and A. Quarteroni. On the coupling of 3D and 1D Navier-Stokes equations for flow problems in compliant vessels. *Comput. Methods Appl. Mech. Engrg.*, 191:561–582, 2001.

[11] L. P. Franca and S. L. Frey. Stabilized finite element methods. II. The incompressible Navier-Stokes equations. *Comput. Methods Appl. Mech. Engrg.*, 99: 209–232, 1992.

[12] T. Tezduyar and S. Sathe. Stabilization parameters in SUPG and PSPG formulations. *J. Comput. Appl. Mech.*, 4:71–88, 2003.

A Fully Implicit Compressible Euler Solver for Atmospheric Flows *

Chao Yang[1,2] and Xiao-Chuan Cai[2]

[1] Institute of Software, Chinese Academy of Sciences, Beijing 100190, P. R. China, yangchao@iscas.ac.cn
[2] Department of Computer Science, University of Colorado at Boulder, Boulder, CO 80309, USA, chao.yang@colorado.edu, cai@cs.colorado.edu

1 Introduction

Numerical methods for global atmospheric modeling have been widely studied in many literatures [5, 7, 9]. It is well-recognized that the global atmospheric flows can be modeled by fully compressible Euler equations with almost no approximations necessary [7]. However, due to the multi-scale nature of the global atmosphere and the high cost of computation, other simplified models have been favorably used in most community codes.

There are two main difficulties in using fully compressible Euler equations in atmospheric flow simulations. One is that the fast waves in the equations lead to very restrictive stability conditions for explicit time-stepping methods; see, e.g., [11]. Another difficulty is that the flow is nearly compressible and the low Mach number results in large numerical dissipation errors in many classical numerical schemes.

To deal with the fast acoustic and inertio-gravity waves in the fully compressible model, we develop a fully implicit method so that the time step size is no longer constrained by the stability condition. And to treat the low-Mach number flow, an improved version of the Advection Upstream Splitting Method (AUSM$^+$-up, [8]) is adapted. This technique has been successfully employed for a shallow water model in [12]. In the fully implicit solver, we use an inexact Newton method to solve the nonlinear system arising at each time step; and the linear Jacobian system for each Newton step is then solved by a Krylov subspace method with an additive Schwarz preconditioner. We show by numerical experiments on a machine with thousands of processors that the parallel Newton-Krylov-Schwarz approach works well for fully compressible atmospheric flows.

* CY was supported in part by NSFC under 61170075 and 91130023, in part by 973 and 863 Programs of China under 2011CB309701 and 2010AA012301. XCC was supported in part by NSF under DMS-0913089 and EAR-0934647.

R. Bank et al. (eds.), *Domain Decomposition Methods in Science and Engineering XX*, Lecture Notes in Computational Science and Engineering 91,
DOI 10.1007/978-3-642-35275-1_81,

2 Governing Equations

Various formulations of the governing equations for mesoscale atmospheric models can be found in, e.g., [6]. In this paper, we focus on the compressible Euler equations by restricting the study on two dimensions (the $x-z$ plane) and omitting the Coriolis terms. The compressible Euler equations for the atmosphere take the following form

$$\frac{\partial Q}{\partial t}+\frac{\partial F}{\partial x}+\frac{\partial G}{\partial z}+S=0,$$

where

$$Q=\begin{pmatrix}\rho\\ \rho u\\ \rho w\\ \rho\theta\end{pmatrix},F=\begin{pmatrix}\rho u\\ \rho u^2+p\\ \rho uw\\ \rho u\theta\end{pmatrix},G=\begin{pmatrix}\rho w\\ \rho wu\\ \rho w^2+p\\ \rho w\theta\end{pmatrix},S=\begin{pmatrix}0\\ 0\\ \rho g\\ 0\end{pmatrix}, \tag{1}$$

where $g=9.80665\,\mathrm{m/s^2}$ is the effective gravity on the surface of the Earth. In the equation, the prognostic variables are the density ρ, the velocity (u,w) and the potential temperature θ of the atmosphere. The system is closed with the equation of state

$$p=p_{00}\left(\frac{\rho R\theta}{p_{00}}\right)^{\gamma},$$

where $p_{00}=1013.25\,\mathrm{hPa}$ is the reference pressure on the surface, $R=287.04\,\mathrm{J}/(\mathrm{kg}\cdot\mathrm{K})$ is the gas constant for dry air and $\gamma=1.4$. For the sake of brevity, we assume the computational domain Ω is a rectangle and the boundary conditions are given in Sect. 5. In some cases, a physical dissipation is added to the left-hand-side of the momentum and velocity equations. The dissipation term is $-\nabla\cdot(\nu\rho\nabla\phi)$ for $\phi=u$, w, and θ.

To recover the hydrostatic solution from the equation, instead of using (1) directly, the following shifted system is often preferred [6, 11]:

$$Q=\begin{pmatrix}\rho'\\ \rho u\\ \rho w\\ (\rho\theta)'\end{pmatrix},F=\begin{pmatrix}\rho u\\ \rho u^2+p'\\ \rho uw\\ \rho u\theta\end{pmatrix},G=\begin{pmatrix}\rho w\\ \rho wu\\ \rho w^2+p'\\ \rho w\theta\end{pmatrix},S=\begin{pmatrix}0\\ 0\\ \rho' g\\ 0\end{pmatrix} \tag{2}$$

where

$$\rho'=\rho-\bar{\rho},\quad p'=p-\bar{p},\quad (\rho\theta)'=\rho\theta-\bar{\rho}\bar{\theta}$$

and the variables with 'bar' satisfy the hydrostatic condition $\frac{\partial\bar{p}}{\partial z}=-\bar{\rho}g$ and $\bar{\theta}$ is obtained from the equation of state. It is clear that the flux Jacobian of the shifted system (2) in each spatial direction is, respectively,

$$\frac{\partial F}{\partial Q}=\begin{pmatrix}0 & 1 & 0 & 0\\ -u^2 & 2u & 0 & c^2/\theta\\ -uw & w & u & 0\\ -u\theta & \theta & 0 & u\end{pmatrix},\quad \frac{\partial G}{\partial Q}=\begin{pmatrix}0 & 0 & 1 & 0\\ -wu & w & u & 0\\ -w^2 & 0 & 2w & c^2/\theta\\ -w\theta & 0 & \theta & w\end{pmatrix},$$

where $c=\sqrt{\gamma p/\rho}$ is the sound speed.

3 Discretizations

Suppose the computational domain is covered by a uniform rectangular $N_x \times N_z$ mesh. Mesh cell $\mathscr{C}_{ij}$ is centered at (x_i, z_j), for $i = 1, \ldots, N_x$ and $j = 1, \ldots, N_z$, with mesh size $\Delta x \times \Delta z$. The solution in cell $\mathscr{C}_{ij}$ at time t is approximated as

$$Q_{ij} \approx \frac{1}{\Delta x \Delta z} \int_{z_j - \Delta z/2}^{z_j + \Delta z/2} \int_{x_i - \Delta x/2}^{x_i + \Delta x/2} Q(x, z, t) dx dz.$$

We employ a cell-centered finite volume method for the spatial discretization of the compressible Euler equations (2). Integrating (2) over $\mathscr{C}_{ij}$ leads to the following semi-discrete system

$$\frac{\partial Q_{i,j}}{\partial t} + \frac{F_{i+1/2,j} - F_{i-1/2,j}}{\Delta x} + \frac{G_{i,j+1/2} - G_{i,j-1/2}}{\Delta z} + S(Q_{i,j}) = 0,$$

where the numerical fluxes of F and G are averaged on the edges of each mesh cell.

To calculate the numerical fluxes on cell edges, we first employ a piecewise linear formulation to reconstruct constant states in both left and right direction, i.e.,

$$Q^-_{i+\frac{1}{2},j} = Q_{ij} + \frac{1}{4}(Q_{i+1,j} - Q_{i-1,j}), \quad Q^+_{i-\frac{1}{2},j} = Q_{ij} - \frac{1}{4}(Q_{i+1,j} - Q_{i-1,j}),$$
$$Q^-_{i,j+\frac{1}{2}} = Q_{ij} + \frac{1}{4}(Q_{i,j+1} - Q_{i,j-1}), \quad Q^+_{i,j-\frac{1}{2}} = Q_{ij} - \frac{1}{4}(Q_{i,j+1} - Q_{i,j-1}).$$

Then we use an improved version of the Advection Upstream Splitting Method (AUSM$^+$-up, [8]) to approximate the numerical fluxes based on the reconstructed states. The basic idea of AUSM$^+$-up scheme is to split the flux into two parts, e.g.,

$$F = F^{(c)} + F^{(p)},$$

where the convective flux $F^{(c)} = \rho u (1, u, w, \theta)^T$ and the pressure flux $F^{(p)} = (0, p', 0, 0)^T$ are estimated separately, both in an upwinded manner. For instance, denote the left and right reconstructed states for the prognostic variables on an edge of a mesh cell as $(\rho_-, u_-, w_-, \theta_-)$ and $(\rho_+, u_+, w_+, \theta_+)$, the pressure flux is approximated by $F^{(p)} \approx (0, \widetilde{p}', 0, 0)^T$, where

$$\widetilde{p}' = \mathscr{P}_5^+(M_-) p'_- + \mathscr{P}_5^-(M_+) p'_+ - (3/2) \mathscr{P}_5^+(M_-) \mathscr{P}_5^-(M_+) \widetilde{\rho}\, \widetilde{c} (u_+ - u_-),$$

and

$$\widetilde{\rho} = (\rho_- + \rho_+)/2, \quad \widetilde{c} = (\sqrt{\gamma p_+ / \rho_+} + \sqrt{\gamma p_- / \rho_-})/2, \quad p'_\pm = p_\pm - \bar{p},$$
$$\mathscr{P}_5^\pm(M) = \begin{cases} (1 \pm \mathrm{sign}(M))/2, & \text{if } |M| \geq 1, \\ \mathscr{M}_2^\pm(M) \left[(\pm 2 - M) \mp 3 M \mathscr{M}_2^\mp(M) \right], & \text{otherwise}, \end{cases}$$
$$\mathscr{M}_2^\pm(M) = (M \pm 1)^2/4, \quad M_\pm = u_\pm / \widetilde{c}.$$

More details can be found in [8].

For the temporal integration, instead of using explicit methods that suffer from severe stability restriction on the time step size, we employ a fully implicit method. Given a semi-discrete system

$$\frac{\partial Q}{\partial t} + \mathscr{L}(Q) = 0,$$

we use the following second-order backward differentiation formula (BDF-2):

$$\frac{1}{2\Delta t}\left(3Q^{(k)} - 4Q^{(k-1)} + Q^{(k-2)}\right) + \mathscr{L}(Q^{(k)}) = 0.$$

Here $Q^{(k)}$ denotes the solution vector Q evaluated at the k-th time step with a fixed time step size Δt. Only at the first time step, a first-order backward Euler method is used.

4 Newton-Krylov-Schwarz Solver

The fully implicit method leads to a large sparse nonlinear algebraic system at each time step. In this study, we use the Newton-Krylov-Schwarz (NKS) algorithm as the nonlinear solver. Given a nonlinear system $\mathscr{F}(X) = 0$, an inexact Newton method is used to solve the system in the outer loop of the NKS approach. Let X_n be the approximate solution for the n-th Newton iterate, we find the next solution X_{n+1} as

$$X_{n+1} = X_n + \lambda_n s_n, \quad n = 0, 1, \ldots$$

where λ_n is the steplength decided by a linesearch procedure and s_n is the Newton correction. We then use the right-preconditioned GMRES (restarted every 30 iterations) method to solve the Jacobian system

$$J_n M^{-1}(M s_n) = -\mathscr{F}(X_n), \quad J_n = \mathscr{F}'(X_n)$$

until the linear residual $r_n = J_n s_n + \mathscr{F}(X_n)$ satisfies

$$\|r_n\| \leq \eta \|\mathscr{F}(X_n)\|,$$

where $\eta > 0$ is the nonlinear forcing term that has been set to be a fixed value $\eta = 1.0 \times 10^{-6}$ in our test. A multi-coloring finite difference method [4] is used to form the Jacobian J_n in the calculation. To achieve uniform residual error at each time step, we use the same adaptive stopping conditions as in [13].

Given the computational domain Ω, we first decompose it into non-overlapping subdomains Ω_k, $k = 1, \ldots, np$, where np is the number of subdomains and also the number of processor cores. Then each subdomain Ω_k is extended to Ω_k^δ within Ω and the number of overlapping mesh layers between subdomains is δ. For the overlapping domain decomposition, a preconditioner M^{-1} is then constructed using the one-level restricted additive Schwarz (RAS, [2]) method defined as follows

$$M^{-1} = \sum_{k=1}^{np} (R_k^0)^T (J_n)_k^{-1} R_k^\delta .$$

Here $(J_n)_k$ is the Jacobian matrix defined on subdomain Ω_k^δ and R_k^δ and $(R_k^0)^T$ are restriction and prolongation operators respectively. Given a solution vector defined on Ω, R_k^δ restricts the vector to the overlapping subdomain Ω_k^δ while $(R_k^0)^T$ prolongates the restricted vector back to the whole domain Ω by putting zeros not only outside Ω_k^δ but also within $\Omega_k^\delta \setminus \Omega_k$. In the implementation of the NKS solver, we use a point-block ordering for both the unknowns and the nonlinear equations, resulting in Jacobian matrices with 4×4-block entries. A point-block version of sparse LU factorization is then used to solve the subdomain problems.

5 Numerical Results

An IBM BlueGene/L supercomputer with 4,096 nodes is used to conduct our numerical tests. Each node of the computer has a dual-core IBM PowerPC 440 processor running at 700 MHz and 512 MB local memory. We implement the NKS algorithm based on the Portable, Extensible Toolkits for Scientific computations (PETSc, [1]) library. In the numerical tests, the overlapping factor in the NKS solver is fixed at $\delta = 2$.

We study a test case describing a rising thermal bubble that is similar to those studied in [3] and [10]. The computational domain is

$$\Omega = \left\{ (x,z) \middle| x \in [-10.0\,\text{km}, 10.0\,\text{km}], z \in [0, 10.0\,\text{km}] \right\},$$

which is assumed to be horizontally periodic with rigid walls (zero normal velocity, i.e., $w = 0$ here) at the bottom and top boundaries. The initial condition for the problem is obtained from a hydrostatic state with $u = w = 0$ and $\bar{\theta} = 300\,\text{K}$ by adding a perturbation

$$\Delta\theta = \begin{cases} 2.0\cos(0.5\pi L)\,\text{K} & \text{if } L \le 1.0, \\ 0.0\,\text{K} & \text{otherwise,} \end{cases}$$

where

$$L = \sqrt{\left(\frac{x - 0.0\,\text{km}}{2.0\,\text{km}}\right)^2 + \left(\frac{z - 2.0\,\text{km}}{2.0\,\text{km}}\right)^2}.$$

A physical dissipation $\nu = 15.0\,\text{m}^2/\text{s}$ is employed in the calculation. The results on a $1,000 \times 500$ mesh using the fully implicit method with $\Delta t = 2.0\,\text{s}$ are provided in Fig. 1. We find that the results are in agreement with those provided in several publications; see, e.g., [3, 10] and [6].

To investigate the performance of the preconditioner, we run a fixed size problem on a $1,920 \times 960$ mesh for 50 time steps with $\Delta t = 2.0\,\text{s}$ by using gradually doubled numbers of processor cores (np). The results on the averaged number of Newton and GMRES iterations per time step are provided in Fig. 2, from where we observe that

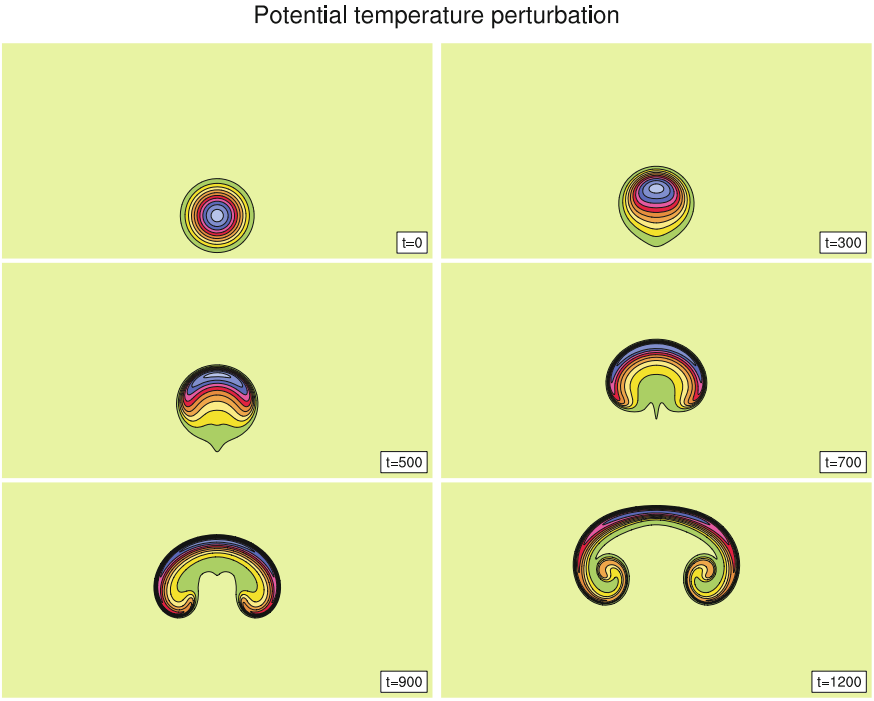


Fig. 1. Contour plots of the potential temperature perturbation (contour interval: 0.2 K)

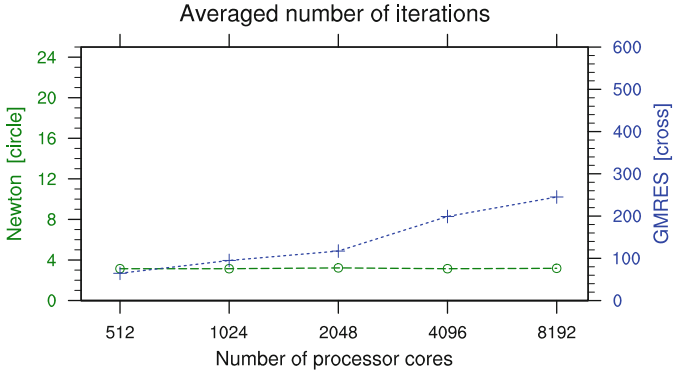


Fig. 2. Averaged numbers of Newton and GMRES iterations per time step

the number of Newton iterations is not sensitive to np but the number of GMRES iterations needed for each time step increases as np increases. The total compute time and the parallel scalability are provided in Fig. 3, which clearly shows that as more processors are used for the fixed size problem, the total compute time is reduced accordingly and the parallel scalability from 512 to 8,192 processor cores is nearly

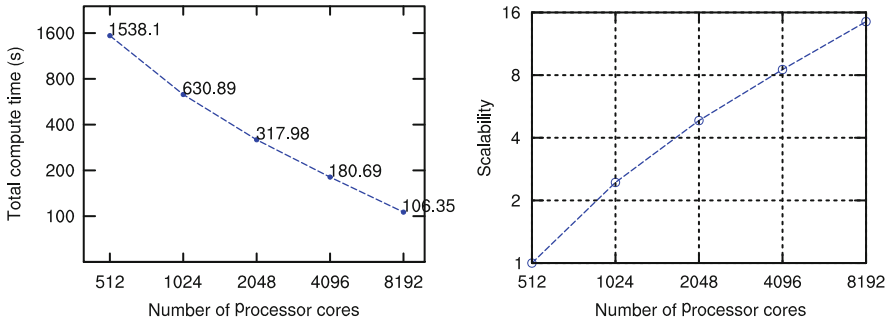


Fig. 3. Total compute time (*left*) and parallel scalability (*right*) results

optimal, with the parallel efficiency reaching 90.38 %. Because of the page limit, we only present a one-level restricted addtive Schwarz method for the compressible Euler problem and only provide some preliminary results in this paper. More advanced algorithms such as multilevel hybrid Schwarz methods will be investigated in a forthcoming paper and more numerical experiments will be carried out in it.

Bibliography

[1] S. Balay, K. Buschelman, V. Eijkhout, W. Gropp, D. Kaushik, M. Knepley, L. C. McInnes, B. Smith, and H. Zhang. *PETSc Users Manual*. Argonne National Laboratory, 2010.

[2] X.-C. Cai and M. Sarkis. A restricted additive Schwarz preconditioner for general sparse linear systems. *SIAM J. Sci. Comput.*, 21:792–797, 1999.

[3] R. L. Carpenter Jr., K. K. Droegemeier, P. R. Woodward, and C. E. Hane. Application of the piecewise parabolic method (PPM) to meteorological modeling. *Mon. Wea. Rev.*, 118:586–612, 1990.

[4] T. F. Coleman and J. J. Moré. Estimation of sparse Jacobian matrices and graph coloring problems. *SIAM J. Numer. Anal.*, 20:187–209, 1983.

[5] K. J. Evans, M. A. Taylor, and J. B. Drake. Accuracy analysis of a spectral element atmospheric model using a fully implicit solution framework. *Mon. Wea. Rev.*, 138:3333–3341, 2010.

[6] F. X. Giraldo and M. Restelli. A study of spectral element and discontinuous Galerkin methods for the Navier-Stokes equations in nonhydrostatic mesoscale atmospheric modeling: Equation sets and test cases. *J. Comput. Phys.*, 227: 3849–3877, 2008.

[7] P. H. Lauritzen, C. Jablonowski, M. A. Taylor, and R. D. Nair, editors. *Numerical Techniques for Global Atmospheric Models*. Springer, 2011.

[8] M.-S. Liou. A sequel to AUSM, part II: AUSM+-up for all speeds. *J. Comput. Phys.*, 214:137–170, 2006.

[9] R. D. Nair, H.-W. Choi, and H. M. Tufo. Computational aspects of a scalable high-order discontinuous Galerkin atmospheric dynamical core. *Comp. Fluids*, 38:309–319, 2009.

[10] A. Robert. Bubble convection experiments with a semi-implicit formulation of the Euler equations. *J. Atmos. Sci.*, 50:1865–1873, 1993.

[11] A. St-Cyr and D. Neckels. A fully implicit Jacobian-free high-order discontinuous Galerkin mesoscale flow solver. In *ICCS 2009, part II, vol. 5545 of Lecture Notes in Computer Science*, pages 243–252. Springer-Verlag, 2009.

[12] P. A. Ullrich, C. Jablonowski, and B. van Leer. High-order finite-volume methods for the shallow-water equations on the sphere. *J. Comput. Phys.*, 229: 6104–6134, 2010.

[13] C. Yang and X.-C. Cai. Parallel multilevel methods for implicit solution of shallow water equations with nonsmooth topography on the cubed-sphere. *J. Comput. Phys.*, 230:2523–2539, 2011.

Editorial Policy

1. Volumes in the following three categories will be published in LNCSE:

i) Research monographs
ii) Tutorials
iii) Conference proceedings

Those considering a book which might be suitable for the series are strongly advised to contact the publisher or the series editors at an early stage.

2. Categories i) and ii). Tutorials are lecture notes typically arising via summer schools or similar events, which are used to teach graduate students. These categories will be emphasized by Lecture Notes in Computational Science and Engineering. **Submissions by interdisciplinary teams of authors are encouraged.** The goal is to report new developments – quickly, informally, and in a way that will make them accessible to non-specialists. In the evaluation of submissions timeliness of the work is an important criterion. Texts should be well-rounded, well-written and reasonably self-contained. In most cases the work will contain results of others as well as those of the author(s). In each case the author(s) should provide sufficient motivation, examples, and applications. In this respect, Ph.D. theses will usually be deemed unsuitable for the Lecture Notes series. Proposals for volumes in these categories should be submitted either to one of the series editors or to Springer-Verlag, Heidelberg, and will be refereed. A provisional judgement on the acceptability of a project can be based on partial information about the work: a detailed outline describing the contents of each chapter, the estimated length, a bibliography, and one or two sample chapters – or a first draft. A final decision whether to accept will rest on an evaluation of the completed work which should include

– at least 100 pages of text;
– a table of contents;
– an informative introduction perhaps with some historical remarks which should be accessible to readers unfamiliar with the topic treated;
– a subject index.

3. Category iii). Conference proceedings will be considered for publication provided that they are both of exceptional interest and devoted to a single topic. One (or more) expert participants will act as the scientific editor(s) of the volume. They select the papers which are suitable for inclusion and have them individually refereed as for a journal. Papers not closely related to the central topic are to be excluded. Organizers should contact the Editor for CSE at Springer at the planning stage, see *Addresses* below.

In exceptional cases some other multi-author-volumes may be considered in this category.

4. Only works in English will be considered. For evaluation purposes, manuscripts may be submitted in print or electronic form, in the latter case, preferably as pdf- or zipped ps-files. Authors are requested to use the LaTeX style files available from Springer at http://www.springer.com/authors/book+authors/helpdesk?SGWID=0-1723113-12-971304-0 (Click on Templates → LaTeX → monographs or contributed books).

For categories ii) and iii) we strongly recommend that all contributions in a volume be written in the same LaTeX version, preferably LaTeX2e. Electronic material can be included if appropriate. Please contact the publisher.

Careful preparation of the manuscripts will help keep production time short besides ensuring satisfactory appearance of the finished book in print and online.

5. The following terms and conditions hold. Categories i), ii) and iii):

Authors receive 50 free copies of their book. No royalty is paid.
Volume editors receive a total of 50 free copies of their volume to be shared with authors, but no royalties.

Authors and volume editors are entitled to a discount of 33.3 % on the price of Springer books purchased for their personal use, if ordering directly from Springer.

6. Commitment to publish is made by letter of intent rather than by signing a formal contract. Springer-Verlag secures the copyright for each volume.

Addresses:

Timothy J. Barth
NASA Ames Research Center
NAS Division
Moffett Field, CA 94035, USA
barth@nas.nasa.gov

Michael Griebel
Institut für Numerische Simulation
der Universität Bonn
Wegelerstr. 6
53115 Bonn, Germany
griebel@ins.uni-bonn.de

David E. Keyes
Mathematical and Computer Sciences
and Engineering
King Abdullah University of Science
and Technology
P.O. Box 55455
Jeddah 21534, Saudi Arabia
david.keyes@kaust.edu.sa

and

Department of Applied Physics
and Applied Mathematics
Columbia University
500 W. 120 th Street
New York, NY 10027, USA
kd2112@columbia.edu

Risto M. Nieminen
Department of Applied Physics
Aalto University School of Science
and Technology
00076 Aalto, Finland
risto.nieminen@aalto.fi

Dirk Roose
Department of Computer Science
Katholieke Universiteit Leuven
Celestijnenlaan 200A
3001 Leuven-Heverlee, Belgium
dirk.roose@cs.kuleuven.be

Tamar Schlick
Department of Chemistry
and Courant Institute
of Mathematical Sciences
New York University
251 Mercer Street
New York, NY 10012, USA
schlick@nyu.edu

Editor for Computational Science
and Engineering at Springer:
Martin Peters
Springer-Verlag
Mathematics Editorial IV
Tiergartenstrasse 17
69121 Heidelberg, Germany
martin.peters@springer.com

Lecture Notes in Computational Science and Engineering

1. D. Funaro, *Spectral Elements for Transport-Dominated Equations.*
2. H.P. Langtangen, *Computational Partial Differential Equations.* Numerical Methods and Diffpack Programming.
3. W. Hackbusch, G. Wittum (eds.), *Multigrid Methods V.*
4. P. Deuflhard, J. Hermans, B. Leimkuhler, A.E. Mark, S. Reich, R.D. Skeel (eds.), *Computational Molecular Dynamics: Challenges, Methods, Ideas.*
5. D. Kröner, M. Ohlberger, C. Rohde (eds.), *An Introduction to Recent Developments in Theory and Numerics for Conservation Laws.*
6. S. Turek, *Efficient Solvers for Incompressible Flow Problems.* An Algorithmic and Computational Approach.
7. R. von Schwerin, ***M**ulti **B**ody **S**ystem **SIM**ulation.* Numerical Methods, Algorithms, and Software.
8. H.-J. Bungartz, F. Durst, C. Zenger (eds.), *High Performance Scientific and Engineering Computing.*
9. T.J. Barth, H. Deconinck (eds.), *High-Order Methods for Computational Physics.*
10. H.P. Langtangen, A.M. Bruaset, E. Quak (eds.), *Advances in Software Tools for Scientific Computing.*
11. B. Cockburn, G.E. Karniadakis, C.-W. Shu (eds.), *Discontinuous Galerkin Methods.* Theory, Computation and Applications.
12. U. van Rienen, *Numerical Methods in Computational Electrodynamics.* Linear Systems in Practical Applications.
13. B. Engquist, L. Johnsson, M. Hammill, F. Short (eds.), *Simulation and Visualization on the Grid.*
14. E. Dick, K. Riemslagh, J. Vierendeels (eds.), *Multigrid Methods VI.*
15. A. Frommer, T. Lippert, B. Medeke, K. Schilling (eds.), *Numerical Challenges in Lattice Quantum Chromodynamics.*
16. J. Lang, *Adaptive Multilevel Solution of Nonlinear Parabolic PDE Systems.* Theory, Algorithm, and Applications.
17. B.I. Wohlmuth, *Discretization Methods and Iterative Solvers Based on Domain Decomposition.*
18. U. van Rienen, M. Günther, D. Hecht (eds.), *Scientific Computing in Electrical Engineering.*
19. I. Babuška, P.G. Ciarlet, T. Miyoshi (eds.), *Mathematical Modeling and Numerical Simulation in Continuum Mechanics.*
20. T.J. Barth, T. Chan, R. Haimes (eds.), *Multiscale and Multiresolution Methods.* Theory and Applications.
21. M. Breuer, F. Durst, C. Zenger (eds.), *High Performance Scientific and Engineering Computing.*
22. K. Urban, *Wavelets in Numerical Simulation.* Problem Adapted Construction and Applications.

23. L.F. Pavarino, A. Toselli (eds.), *Recent Developments in Domain Decomposition Methods.*

24. T. Schlick, H.H. Gan (eds.), *Computational Methods for Macromolecules: Challenges and Applications.*

25. T.J. Barth, H. Deconinck (eds.), *Error Estimation and Adaptive Discretization Methods in Computational Fluid Dynamics*.

26. M. Griebel, M.A. Schweitzer (eds.), *Meshfree Methods for Partial Differential Equations*.

27. S. Müller, *Adaptive Multiscale Schemes for Conservation Laws*.

28. C. Carstensen, S. Funken, W. Hackbusch, R.H.W. Hoppe, P. Monk (eds.), *Computational Electromagnetics*.

29. M.A. Schweitzer, *A Parallel Multilevel Partition of Unity Method for Elliptic Partial Differential Equations*.

30. T. Biegler, O. Ghattas, M. Heinkenschloss, B. van Bloemen Waanders (eds.), *Large-Scale PDE-Constrained Optimization*.

31. M. Ainsworth, P. Davies, D. Duncan, P. Martin, B. Rynne (eds.), *Topics in Computational Wave Propagation*. Direct and Inverse Problems.

32. H. Emmerich, B. Nestler, M. Schreckenberg (eds.), *Interface and Transport Dynamics.* Computational Modelling.

33. H.P. Langtangen, A. Tveito (eds.), *Advanced Topics in Computational Partial Differential Equations.* Numerical Methods and Diffpack Programming.

34. V. John, *Large Eddy Simulation of Turbulent Incompressible Flows.* Analytical and Numerical Results for a Class of LES Models.

35. E. Bänsch (ed.), *Challenges in Scientific Computing - CISC 2002.*

36. B.N. Khoromskij, G. Wittum, *Numerical Solution of Elliptic Differential Equations by Reduction to the Interface.*

37. A. Iske, *Multiresolution Methods in Scattered Data Modelling.*

38. S.-I. Niculescu, K. Gu (eds.), *Advances in Time-Delay Systems.*

39. S. Attinger, P. Koumoutsakos (eds.), *Multiscale Modelling and Simulation.*

40. R. Kornhuber, R. Hoppe, J. Périaux, O. Pironneau, O. Wildlund, J. Xu (eds.), *Domain Decomposition Methods in Science and Engineering.*

41. T. Plewa, T. Linde, V.G. Weirs (eds.), *Adaptive Mesh Refinement – Theory and Applications.*

42. A. Schmidt, K.G. Siebert, *Design of Adaptive Finite Element Software.* The Finite Element Toolbox ALBERTA.

43. M. Griebel, M.A. Schweitzer (eds.), *Meshfree Methods for Partial Differential Equations II.*

44. B. Engquist, P. Lötstedt, O. Runborg (eds.), *Multiscale Methods in Science and Engineering.*

45. P. Benner, V. Mehrmann, D.C. Sorensen (eds.), *Dimension Reduction of Large-Scale Systems.*

46. D. Kressner, *Numerical Methods for General and Structured Eigenvalue Problems.*

47. A. Boriçi, A. Frommer, B. Joó, A. Kennedy, B. Pendleton (eds.), *QCD and Numerical Analysis III*.

48. F. Graziani (ed.), *Computational Methods in Transport.*

49. B. Leimkuhler, C. Chipot, R. Elber, A. Laaksonen, A. Mark, T. Schlick, C. Schütte, R. Skeel (eds.), *New Algorithms for Macromolecular Simulation.*

50. M. Bücker, G. Corliss, P. Hovland, U. Naumann, B. Norris (eds.), *Automatic Differentiation: Applications, Theory, and Implementations.*

51. A.M. Bruaset, A. Tveito (eds.), *Numerical Solution of Partial Differential Equations on Parallel Computers.*

52. K.H. Hoffmann, A. Meyer (eds.), *Parallel Algorithms and Cluster Computing.*

53. H.-J. Bungartz, M. Schäfer (eds.), *Fluid-Structure Interaction.*

54. J. Behrens, *Adaptive Atmospheric Modeling.*

55. O. Widlund, D. Keyes (eds.), *Domain Decomposition Methods in Science and Engineering XVI.*

56. S. Kassinos, C. Langer, G. Iaccarino, P. Moin (eds.), *Complex Effects in Large Eddy Simulations.*

57. M. Griebel, M.A Schweitzer (eds.), *Meshfree Methods for Partial Differential Equations III.*

58. A.N. Gorban, B. Kégl, D.C. Wunsch, A. Zinovyev (eds.), *Principal Manifolds for Data Visualization and Dimension Reduction.*

59. H. Ammari (ed.), *Modeling and Computations in Electromagnetics: A Volume Dedicated to Jean-Claude Nédélec.*

60. U. Langer, M. Discacciati, D. Keyes, O. Widlund, W. Zulehner (eds.), *Domain Decomposition Methods in Science and Engineering XVII.*

61. T. Mathew, *Domain Decomposition Methods for the Numerical Solution of Partial Differential Equations.*

62. F. Graziani (ed.), *Computational Methods in Transport: Verification and Validation.*

63. M. Bebendorf, *Hierarchical Matrices.* A Means to Efficiently Solve Elliptic Boundary Value Problems.

64. C.H. Bischof, H.M. Bücker, P. Hovland, U. Naumann, J. Utke (eds.), *Advances in Automatic Differentiation.*

65. M. Griebel, M.A. Schweitzer (eds.), *Meshfree Methods for Partial Differential Equations IV*.

66. B. Engquist, P. Lötstedt, O. Runborg (eds.), *Multiscale Modeling and Simulation in Science.*

67. I.H. Tuncer, Ü. Gülcat, D.R. Emerson, K. Matsuno (eds.), *Parallel Computational Fluid Dynamics 2007.*

68. S. Yip, T. Diaz de la Rubia (eds.), *Scientific Modeling and Simulations*.

69. A. Hegarty, N. Kopteva, E. O'Riordan, M. Stynes (eds.), *BAIL* 2008 – *Boundary and Interior Layers.*

70. M. Bercovier, M.J. Gander, R. Kornhuber, O. Widlund (eds.), *Domain Decomposition Methods in Science and Engineering XVIII.*

71. B. Koren, C. Vuik (eds.), *Advanced Computational Methods in Science and Engineering.*

72. M. Peters (ed.), *Computational Fluid Dynamics for Sport Simulation.*

73. H.-J. Bungartz, M. Mehl, M. Schäfer (eds.), *Fluid Structure Interaction II - Modelling, Simulation, Optimization.*

74. D. Tromeur-Dervout, G. Brenner, D.R. Emerson, J. Erhel (eds.), *Parallel Computational Fluid Dynamics 2008.*

75. A.N. Gorban, D. Roose (eds.), *Coping with Complexity: Model Reduction and Data Analysis.*

76. J.S. Hesthaven, E.M. Rønquist (eds.), *Spectral and High Order Methods for Partial Differential Equations.*

77. M. Holtz, *Sparse Grid Quadrature in High Dimensions with Applications in Finance and Insurance*.

78. Y. Huang, R. Kornhuber, O.Widlund, J. Xu (eds.), *Domain Decomposition Methods in Science and Engineering XIX*.

79. M. Griebel, M.A. Schweitzer (eds.), *Meshfree Methods for Partial Differential Equations V*.

80. P.H. Lauritzen, C. Jablonowski, M.A. Taylor, R.D. Nair (eds.), *Numerical Techniques for Global Atmospheric Models*.

81. C. Clavero, J.L. Gracia, F.J. Lisbona (eds.), *BAIL 2010 – Boundary and Interior Layers, Computational and Asymptotic Methods*.

82. B. Engquist, O. Runborg, Y.R. Tsai (eds.), *Numerical Analysis and Multiscale Computations*.

83. I.G. Graham, T.Y. Hou, O. Lakkis, R. Scheichl (eds.), *Numerical Analysis of Multiscale Problems*.

84. A. Logg, K.-A. Mardal, G. Wells (eds.), *Automated Solution of Differential Equations by the Finite Element Method*.

85. J. Blowey, M. Jensen (eds.), *Frontiers in Numerical Analysis - Durham 2010*.

86. O. Kolditz, U.-J. Gorke, H. Shao, W. Wang (eds.), *Thermo-Hydro-Mechanical-Chemical Processes in Fractured Porous Media - Benchmarks and Examples*.

87. S. Forth, P. Hovland, E. Phipps, J. Utke, A. Walther (eds.), *Recent Advances in Algorithmic Differentiation*.

88. J. Garcke, M. Griebel (eds.), *Sparse Grids and Applications*.

89. M. Griebel, M.A. Schweitzer (eds.), *Meshfree Methods for Partial Differential Equations VI*.

90. C. Pechstein, *Finite and Boundary Element Tearing and Interconnecting Solvers for Multiscale Problems*.

91. R. Bank, M. Holst, O. Widlund, J. Xu (eds.), *Domain Decomposition Methods in Science and Engineering XX*.

For further information on these books please have a look at our mathematics catalogue at the following URL: www.springer.com/series/3527

Monographs in Computational Science and Engineering

1. J. Sundnes, G.T. Lines, X. Cai, B.F. Nielsen, K.-A. Mardal, A. Tveito, *Computing the Electrical Activity in the Heart.*

For further information on this book, please have a look at our mathematics catalogue at the following URL: www.springer.com/series/7417

Texts in Computational Science and Engineering

1. H. P. Langtangen, *Computational Partial Differential Equations.* Numerical Methods and Diffpack Programming. 2nd Edition
2. A. Quarteroni, F. Saleri, P. Gervasio, *Scientific Computing with MATLAB and Octave.* 3rd Edition
3. H. P. Langtangen, *Python Scripting for Computational Science*. 3rd Edition
4. H. Gardner, G. Manduchi, *Design Patterns for e-Science*.
5. M. Griebel, S. Knapek, G. Zumbusch, *Numerical Simulation in Molecular Dynamics*.
6. H. P. Langtangen, *A Primer on Scientific Programming with Python.* 3rd Edition
7. A. Tveito, H. P. Langtangen, B. F. Nielsen, X. Cai, *Elements of Scientific Computing.*
8. B. Gustafsson, *Fundamentals of Scientific Computing.*
9. M. Bader, *Space-Filling Curves.*
10. M. Larson, F. Bengzon, *The Finite Element Method: Theory, Implementation and Applications.*

For further information on these books please have a look at our mathematics catalogue at the following URL: www.springer.com/series/5151