Numerical Algorithms (2024) 96:719–738
https://doi.org/10.1007/s11075-023-01663-1

ORIGINAL PAPER

Rates of robust superlinear convergence of preconditioned Krylov methods for elliptic FEM problems

S. J. Castillo[1] · **J. Karátson**[2,3]

Received: 10 March 2023 / Accepted: 10 September 2023 / Published online: 7 October 2023

Abstract
This paper considers the iterative solution of finite element discretizations of second-order elliptic boundary value problems. Mesh independent estimations are given for the rate of superlinear convergence of preconditioned Krylov methods, involving the connection between the convergence rate and the Lebesgue exponent of the data. Numerical examples demonstrate the theoretical results.

Keywords Elliptic boundary value problems · Finite element method · Preconditioned Krylov iterations · Superlinear convergence

1 Introduction

The preconditioned conjugate gradient (PCG) method is a widespread way for the iterative solution of discretized elliptic partial differential equations. It can be efficiently coupled with multigrid methods and, under certain conditions, operator preconditioning can provide mesh-independent convergence. Hence, its convergence has been widely analyzed, see, e.g., [1, 3] and references therein.

In particular, superlinear convergence is often a characteristic second stage in the convergence history: this notion expresses, roughly speaking, that the number of iterations required to achieve a new correct digit will be decreasing in course of the iteration. This phenomenon is also favorable when the PCG method is used as an inner iteration for an outer process. Such results were already obtained in [11, 15] on operator level.

✉ J. Karátson
kajkaat@caesar.elte.hu

S. J. Castillo
sebastian.castillo@yachaytech.edu.ec

1 Department of Applied Analysis, Eötvös Loránd University, Budapest, Hungary

2 Department of Applied Analysis & HUN-REN-ELTE Numerical Analysis and Large Networks Research Group, Eötvös Loránd University, Budapest, Hungary

3 Department of Analysis and Operations Research, Budapest University of Technology and Economics, Budapest, Hungary

This paper considers some types of second-order elliptic boundary value problems with variable zeroth order coefficients and their finite element discretizations. Our goal is to find relevant estimations for the rate of superlinear convergence of the PCG method for this type of problem; furthermore, we are interested in robust, that is, mesh independent rates, which can be given independently of the finite element mesh size. This means that the favorable behavior does not deteriorate as the mesh is refined.

This mesh-independence property of superlinear convergence was studied in various joint papers of the second author, see, e.g., [2] for a general result, [3] for a survey in this journal, and [5] for some recent applications. The starting point of the present paper is [10], where a superlinear rate was found in a particular situation with continuous zeroth order coefficient. Our goal is to extend this result to a family of estimations for general zeroth order ("linearized reaction") coefficients, that is, which are unbounded, and belong to some Lebesgue space. Furthermore, we would like to explore the connection between the convergence rate and the Lebesgue exponent. A practical motivation for such situations is, among other things, the Newton linearization arising in reaction-diffusion models where the nonlinear rate of reaction is typically of polynomial order, thus leading to linearized coefficients with given Lebesgue exponent.

We present eigenvalue-based estimations of the rate of superlinear convergence for such problems, first for single equations, then we show that similar estimations can be obtained in the case of proper systems of PDEs, involving GMRES in the nonsymmetric case. Finally, some numerical examples are shown, which properly demonstrate our theoretical results.

2 Theoretical background

2.1 The abstract problem and its discretization

Let H be a real Hilbert space and let us consider a linear operator equation

$$Au = g \tag{1}$$

with some $g \in H$, under the following

Assumption 2.1

(i) The operator A is decomposed as

$$A = S + Q \tag{2}$$

where S is a symmetric operator in H with dense domain D and Q is a compact self-adjoint operator defined on the domain H.

(ii) There exists $k > 0$ such that $\langle Su, u\rangle \geq k\|u\|^2$ $(\forall u \in D)$.

(iii) $\langle Qu, u\rangle \geq 0$ $(\forall u \in H)$.

We recall that the energy space H_S is the completion of D under the *energy inner product*

$$\langle u, v\rangle_S = \langle Su, v\rangle, \tag{3}$$

and the corresponding norm is denoted by $\|\cdot\|_S$. Assumption (ii) implies $H_S \subset H$. Then, there exists a unique bounded linear operator, denoted by $Q_S : H_S \mapsto H_S$, such that

$$\langle Q_S u, v\rangle_S = \langle Qu, v\rangle \qquad (\forall u, v \in H_S).$$

We replace (1) by its formally preconditioned form

$$Bu \equiv S^{-1}Au = S^{-1}g,$$

that is, $(I + S^{-1}Q)u = S^{-1}g$ in H_S. This gives the weak formulation

$$\langle (I + Q_S)u, v\rangle_S = \langle g, v\rangle \qquad (\forall v \in H_S). \tag{4}$$

Since by assumption (iii) the bilinear form on the left is coercive on H_S, by the *Lax-Milgram theorem*, there exists a unique solution $u \in H_S$ of (4).

Now (4) is solved numerically using a *Galerkin discretization*. Consider a given finite-dimensional subspace $V = \text{span}\{\varphi_1, \ldots, \varphi_n\} \subset H_S$, and let

$$\mathbf{S}_h = \{\langle \varphi_i, \varphi_j\rangle_S\}_{i,j=1}^n \quad \text{and } \mathbf{Q}_h = \{\langle Q\varphi_i, \varphi_j\rangle\}_{i,j=1}^n$$

the *Gram matrices* corresponding to S and Q. We look for the numerical solution $u_V \in V$ of (4) in V, i.e., for which

$$\langle (I + Q_S)u_V, v\rangle_S = \langle g, v\rangle \qquad (\forall v \in V). \tag{5}$$

Then, $u_V = \sum_{j=1}^n c_j\varphi_j$, where $\mathbf{c} = (c_1, \ldots, c_n) \in \mathbb{R}^n$ is the solution of the system

$$(\mathbf{S}_h + \mathbf{Q}_h)\mathbf{c} = \mathbf{b} \tag{6}$$

with $\mathbf{b} = \{\langle g, \varphi_j\rangle\}_{j=1}^n$. The matrix $\mathbf{A}_h := \mathbf{S}_h + \mathbf{Q}_h$ is SPD.

By using matrix $\mathbf{S}_h$ as the preconditioner for the system (6), we shall work with the preconditioned system

$$(\mathbf{I} + \mathbf{S}_h^{-1}\mathbf{Q}_h)\mathbf{c} = \tilde{\mathbf{b}}, \tag{7}$$

where $\mathbf{I}$ is the identity matrix in $\mathbb{R}^n$ and $\tilde{\mathbf{b}} = \mathbf{S}_h^{-1}\mathbf{b}$. We apply the CGM for the solution of this system.

2.2 The preconditioned conjugate gradient method and superlinear convergence

Let us consider a general linear system $\mathbf{A}_h\mathbf{u} = \mathbf{g}$ and its preconditioned form

$$\mathbf{B}_h\mathbf{u} = \tilde{\mathbf{g}}, \tag{8}$$

where $\mathbf{B}_h = \mathbf{S}_h^{-1}\mathbf{A}$ and $\tilde{\mathbf{g}} = \mathbf{S}_h^{-1}\mathbf{g}$. The preconditioner $\mathbf{S}_h$ induces the energy inner product $\langle \mathbf{c}, \mathbf{d}\rangle_{\mathbf{S}_h} := \mathbf{S}_h\,\mathbf{c}\cdot\mathbf{d}$, where $\cdot$ denotes the standard Euclidean inner product.

Then, the PCG method is given by the following algorithm. Let $\mathbf{u}_0$ be arbitrary, $\boldsymbol{\rho}_0 = \mathbf{A}_h\mathbf{u}_0 - \mathbf{g}$, $\mathbf{S}_h\mathbf{p}_0 = \boldsymbol{\rho}_0$, $\mathbf{r}_0 = \boldsymbol{\rho}_0$ and for $k \in \mathbb{N}$

$$\begin{cases} \mathbf{u}_{k+1} = \mathbf{u}_k + \alpha_k \mathbf{p}_k, \\ \mathbf{r}_{k+1} = \mathbf{r}_k + \alpha_k \mathbf{S}_h^{-1}\mathbf{A}_h\mathbf{p}_k, \\ \mathbf{p}_{k+1} = \mathbf{r}_{k+1} + \beta_k \mathbf{p}_k \end{cases}$$

with

$$\alpha_k = \frac{-\|\mathbf{r}_k\|^2_{\mathbf{S}_h}}{\langle \mathbf{A}_h\mathbf{p}_k, \mathbf{p}_k\rangle}, \quad \beta_k = \frac{\|\mathbf{r}_{k+1}\|^2_{\mathbf{S}_h}}{\|\mathbf{r}_k\|^2_{\mathbf{S}_h}}.$$

In fact, the vector $\mathbf{z}_k := \mathbf{S}_h^{-1}\mathbf{A}_h\mathbf{p}_k$ is computed by solving the auxiliary problem

$$\mathbf{S}_h\mathbf{z}_k = \mathbf{A}_h\mathbf{p}_k\,.$$

Moreover, setting $\mathbf{w}_k = \mathbf{z}_k - \mathbf{p}_k$, this problem is equivalent to

$$\begin{cases} \mathbf{S}_h\mathbf{w}_k = \mathbf{Q}_h\mathbf{p}_k, \\ \mathbf{z}_k = \mathbf{w}_k + \mathbf{p}_k\,. \end{cases} \tag{9}$$

We are interested in the superlinear convergence rates for the CGM, and now recall the corresponding well-known estimation. Let $\mathbf{A}_h = \mathbf{S}_h + \mathbf{Q}_h$. Then, $\mathbf{B}_h$ in (8) has the compact perturbation form $\mathbf{B}_h = \mathbf{I}_h + \mathbf{E}_h$ with $\mathbf{E}_h := \mathbf{S}_h^{-1}\mathbf{Q}_h$. Let us order the eigenvalues of the latter according to $|\lambda_1(\mathbf{S}_h^{-1}\mathbf{Q}_h)| \geq |\lambda_2(\mathbf{S}_h^{-1}\mathbf{Q}_h)| \geq \cdots \geq |\lambda_n(\mathbf{S}_h^{-1}\mathbf{Q}_h)|$. Then, the *error vectors* $\mathbf{e}_k := \mathbf{c}_k - \mathbf{c}$ are measured by $\langle \mathbf{B}_h\mathbf{e}_k, \mathbf{e}_k\rangle^{1/2}_{\mathbf{S}_h} = \left\langle \mathbf{S}_h^{-1}\mathbf{A}_h\mathbf{e}_k, \mathbf{e}_k\right\rangle^{1/2}_{\mathbf{S}_h} = \langle \mathbf{A}_h\mathbf{e}_k, \mathbf{e}_k\rangle^{1/2} = \|\mathbf{e}_k\|_{\mathbf{A}_h}$, and they are known to satisfy

$$\left(\frac{\|\mathbf{e}_k\|_{\mathbf{A}_h}}{\|\mathbf{e}_0\|_{\mathbf{A}_h}}\right)^{1/k} \leq \frac{2\|\mathbf{B}_h^{-1}\|_{\mathbf{S}_h}}{k}\sum_{j=1}^{k}|\lambda_j(\mathbf{S}_h^{-1}\mathbf{Q}_h)| \qquad (k = 1, 2, \ldots, n). \tag{10}$$

This follows, e.g., from formula (13.13) in [1], see also (2.16) in [3].

For the discretized problem described in subsection 2.1, the following result allows us to give a convergence rate for the upper bound of (10) through the eigenvalues of the operator Q_S. This is a modification of Theorem 1 in [10] where the square of eigenvalues was considered.

Lemma 1 *Let assumptions 2.1 hold. Then, for any* $k = 1, 2, \ldots, n$

$$\sum_{j=1}^{k}|\lambda_j(\mathbf{S}_h^{-1}\mathbf{Q}_h)| \leq \sum_{j=1}^{k}\lambda_j(Q_S). \tag{11}$$

Proof We have in fact

$$\sum_{j=1}^{k} \sigma_j(\mathbf{S}_h^{-1}\mathbf{Q}_h) \le \sum_{j=1}^{k} \sigma_j(Q_S), \tag{12}$$

where the σ_j denote the singular values of the given matrix or operator, see [4]. Now both the matrix $\mathbf{S}_h^{-1}\mathbf{Q}_h$ (with respect to the $\mathbf{S}_h$-inner product) and the operator Q_S (in H_S) are self-adjoint, hence their singular values coincide with the modulus of the eigenvalues. Since Q_S is a positive operator from assumption (iii), the modulus can be omitted. □

An immediate consequence of this lemma is the following mesh-independent bound.

Corollary 1 *For any* $k = 1, 2, \ldots, n$

$$\left(\frac{\|e_k\|_{\mathbf{A}_h}}{\|e_0\|_{\mathbf{A}_h}}\right)^{1/k} \le \frac{2\|B^{-1}\|_S}{k} \sum_{j=1}^{k} \lambda_j(Q_S) \qquad (k = 1, 2, \ldots, n). \tag{13}$$

Proof By [2, Prop. 4.1], we are able to estimate $\|\mathbf{B}_h^{-1}\|_{\mathbf{S}_h} \le \|B^{-1}\|_S$. This, together with (10) and (11), completes the proof. □

Since $|\lambda_1(Q_S)| \ge |\lambda_2(Q_S)| \ge \cdots \ge 0$ and the eigenvalues tend to 0, the convergence factor is less than 1 for k sufficiently large. Hence, the upper bound decreases as $k \to \infty$ and we obtain superlinear convergence rate.

3 Estimation of the rates of superlinear convergence

We present the rate estimates in the following stages. First, we develop the results in detail for single equations. The studied preconditioners have the advantage that the original PDE is reduced to simpler PDEs whose discretizations can be solved by proper optimal fast solvers. Then, we extend the estimates for systems of PDEs, first for the symmetric and then for the nonsymmetric case. This situation shows the real strength of the idea of preconditioning operators, since one can reduce large coupled systems of PDEs to independent single PDEs, hence the numerical solution of the latter can be parallelized. In each case, we provide an estimation of the rate of mesh-independent superlinear convergence such that the dependence of the rate on the integrability exponent of the reaction coefficient is determined.

3.1 Elliptic equations

Let $d \ge 2$ and $\Omega \subset \mathbb{R}^d$ be a bounded domain. We consider the elliptic problem

$$\begin{cases} -\mathrm{div}(G\nabla u) + \eta u = g, \\ u|_{\partial\Omega} = 0 \end{cases} \tag{14}$$

in the following situation.

Assumption 3.1

(i) The symmetric matrix-valued function $G \in L^\infty(\overline{\Omega}, \mathbb{R}^d \times \mathbb{R}^d)$ satisfies

$$G(x)\xi \cdot \xi \geq m|\xi|^2 \qquad (\forall \xi \in \mathbb{R}^d)$$

for some $m > 0$ independent of ξ.

(ii) We have $\eta \geq 0$; furthermore, there exists $2 < p < \frac{2d}{d-2}$ such that

$$\eta \in L^{p/(p-2)}(\Omega). \tag{15}$$

(iii) $\partial\Omega$ is a Lipschitz boundary.

(iv) $g \in L^2(\Omega)$.

Then, problem (14) has a unique weak solution in $H_0^1(\Omega)$. The relevance of the condition on p in (ii) is that the continuous embedding $H_0^1(\Omega) \subset L^p(\Omega)$ holds, which ensures the boundedness of the corresponding bilinear form.

In practice, we are mostly interested in the case when the principal part has constant or separable coefficients, whereas $\eta = \eta(x)$ is a general variable (i.e., nonconstant) coefficient. In this case, the principal part will be an efficient preconditioning operator, see Remark 1 for background and extensions.

Let $V_h \subset H_0^1(\Omega)$ be a given FEM subspace. We look for the numerical solution u_h of (14) in V_h:

$$\int_\Omega (G\nabla u_h \cdot \nabla v + \eta u_h v) = \int_\Omega g v \qquad (\forall v \in V_h). \tag{16}$$

The corresponding linear algebraic system has the form

$$(\mathbf{G}_h + \mathbf{D}_h)\mathbf{c} = \mathbf{g}_h,$$

where $\mathbf{G}_h$ and $\mathbf{D}_h$ are the corresponding weighted stiffness and mass matrices, respectively. We apply the matrix $\mathbf{G}_h$ as preconditioner, thus the preconditioned form of (16) is given by

$$(\mathbf{I}_h + \mathbf{G}_h^{-1}\mathbf{D}_h)\mathbf{c} = \tilde{\mathbf{g}}_h \tag{17}$$

with $\tilde{\mathbf{g}}_h = \mathbf{G}_h^{-1}\mathbf{g}_h$. Then, we apply the CGM to (17). The auxiliary systems with $\mathbf{G}_h$ can be solved efficiently with fast solvers, see Remark 1.

Such equations for $\eta \in C(\overline{\Omega})$ were considered in [10]. That was a rather restrictive assumption, see also Remark 2 for the motivations of the more general case (15).

Theorem 1 *Let Assumptions 3.1 hold. Then, there exists $C > 0$ such that for all $k \in \mathbb{N}$*

$$\left(\frac{\|e_k\|_{\mathbf{A}_h}}{\|e_0\|_{\mathbf{A}_h}}\right)^{\frac{1}{k}} \leq Ck^{-\alpha}, \tag{18}$$

where $\alpha = \frac{1}{d} - \frac{1}{2} + \frac{1}{p}$.

Proof Let us consider the real Hilbert space $L^2(\Omega)$ endowed with the usual inner product. Let $D = H^1_{0,G} := \{u \in H^1_0(\Omega) \cap H^2(\Omega) : \ G\nabla u \in H(\mathrm{div}, \Omega)\}$. We define the operators

$$Su \equiv -\mathrm{div}(G\nabla u) \quad (u \in D) \quad \text{and} \quad Qu \equiv \eta u \quad (u \in H^1_0(\Omega)). \tag{19}$$

Then,

$$\langle Su, u\rangle_{L^2} \geq m \int_\Omega |\nabla u|^2 \geq mC_\Omega \int_\Omega u^2 \qquad (\forall u \in D), \tag{20}$$

where C_Ω is the Poincaré–Friedrichs constant and m is the lower spectral bound of G given by assumption (i). Hence the energy space H_S is a well-defined Hilbert space with $\langle u, v\rangle_S = \int_\Omega G\nabla u \cdot \nabla v$. It is easy to see that $H_S = H^1_0(\Omega)$ and that the following inequality holds:

$$\sqrt{m}\|u\|_{H^1_0(\Omega)} \leq \|u\|_{H_S} \qquad (\forall u \in H_S). \tag{21}$$

Since $p < \frac{2d}{d-2}$, the embedding $\mathcal{I} : H^1_0(\Omega) \to L^p(\Omega)$ is compact, in particular, there exists $\hat{c} > 0$ such that for all $u \in H^1_0(\Omega)$

$$\|u\|_{L^p(\Omega)} \leq \hat{c}\|u\|_{H^1_0(\Omega)}. \tag{22}$$

Then,

$$\begin{aligned}
\|Q_S v\|_{H_S} = \sup_{\|u\|_{H_S}=1} |\langle Q_S v, u\rangle_S| &= \sup_{\|u\|_{H_S}=1} \langle Qv, u\rangle \\
&= \sup_{\|u\|_{H_S}=1} \int_\Omega \eta v u \\
&\leq \sup_{\|u\|_{H_S}=1} \left(\int_\Omega |\eta|^{\frac{p}{p-2}}\right)^{\frac{p-2}{p}} \left(\int_\Omega |v|^p\right)^{\frac{1}{p}} \left(\int_\Omega |u|^p\right)^{\frac{1}{p}} \\
&\leq \hat{c} \sup_{\|u\|_{H_S}=1} \|\eta\|_{L^{p/(p-2)}(\Omega)} \|v\|_{L^p(\Omega)} \|u\|_{H^1_0(\Omega)} \\
&\leq \frac{\hat{c}}{\sqrt{m}} \sup_{\|u\|_{H_S}=1} \|\eta\|_{L^{p/(p-2)}(\Omega)} \|v\|_{L^p(\Omega)} \|u\|_{H_S} \\
&= \frac{\hat{c}M}{\sqrt{m}} \|v\|_{L^p(\Omega)},
\end{aligned} \tag{23}$$

where $M = \|\eta\|_{L^{p/(p-2)}(\Omega)}$. Here, we applied the extension of Hölder's inequality ([6, Th. 4.6]) with

$$1 = \frac{1}{p} + \frac{1}{p} + \frac{p-2}{p}.$$

Hence, Q_S is compact in H_S. Altogether, Q_S is a compact self-adjoint operator in H_S, hence, by [9, Ch.6, Th.1.5], we have the following characterization of the eigenvalues

of Q_S:

$$\forall n \in \mathbb{N}: \quad \lambda_n(Q_S) = \min\{\|Q_S - L_{n-1}\| : \ L_{n-1} \in \mathcal{L}(H_S),\ \operatorname{rank}(L_{n-1}) \le n-1\}. \tag{24}$$

By taking the minimum over a smaller subset of finite rank operators, we obtain

$$\lambda_n(Q_S) \le \min\{\|Q_S - Q_S L_{n-1}\| : \ L_{n-1} \in \mathcal{L}(H_S),\ \operatorname{rank}(L_{n-1}) \le n-1\}. \tag{25}$$

Now, by (23) and (21) we get

$$\begin{aligned}
\|Q_S - Q_S L_{n-1}\| &= \sup_{u \in H_S} \frac{\|(Q_S - Q_S L_{n-1})u\|_{H_S}}{\|u\|_{H_S}} \\
&= \sup_{u \in H_S} \frac{\|Q_S(u - L_{n-1}u)\|_{H_S}}{\|u\|_{H_S}} \\
&\le \frac{\hat{c}M}{\sqrt{m}} \sup_{u \in H_S} \frac{\|u - L_{n-1}u\|_{L^p(\Omega)}}{\|u\|_{H_S}} \\
&\le \frac{\hat{c}M}{\sqrt{m}\sqrt{m}} \sup_{u \in H_0^1(\Omega)} \frac{\|u - L_{n-1}u\|_{L^p(\Omega)}}{\|u\|_{H_0^1(\Omega)}}.
\end{aligned}$$

This, together with (25) yields

$$\begin{aligned}
\lambda_n(Q_S) &\le \frac{\hat{c}M}{m} \min\{\|\mathcal{I} - L_{n-1}\| : \ L_{n-1} \in \mathcal{L}(H_0^1(\Omega), L^p(\Omega)),\ \operatorname{rank}(L_{n-1}) \le n-1\} \\
&:= \frac{\hat{c}M}{m} a_n(\mathcal{I}),
\end{aligned} \tag{26}$$

where $a_n(\mathcal{I})$ denotes the approximation numbers of the compact embedding $\mathcal{I}\colon H_0^1(\Omega) \mapsto L^p(\Omega)$, see [14]. Furthermore, we have the estimation from [8]:

$$a_n(\mathcal{I}) \le \hat{C} n^{-\alpha}, \qquad \text{where} \quad \alpha = \frac{1}{d} - \frac{1}{2} + \frac{1}{p} \tag{27}$$

for some constant $\hat{C} > 0$. Therefore, we arrive at the inequality

$$\lambda_n(Q_S) \le \frac{\hat{C}\hat{c}M}{m} n^{-\alpha}.$$

Now, taking the arithmetic mean on both sides and estimating the sum from above by an integral, we obtain

$$\frac{1}{k}\sum_{n=1}^{k} \lambda_n(Q_S) \le \frac{\hat{C}\hat{c}M}{m} \frac{1}{k} \left(1 + \int_1^k \frac{1}{x^\alpha}\, dx\right) \le \frac{\hat{C}\hat{c}M}{m(1-\alpha)} \frac{1}{k^\alpha}. \tag{28}$$

Then, by (13), we conclude. □

Remark 1 The PCG method requires the solution of auxiliary problems $\mathbf{S}w_k = \mathbf{Q}p_k$, see (9). In the main case, when the principal part has constant or separable coefficients, these problems can be solved easily with fast solvers due to the special structure of the operator $Su \equiv -\mathrm{div}(G\nabla u)$ (in particular, when $Su = -\Delta u$), see, e.g., [7], [12].

Moreover, to generalize the above, one may also incorporate a constant lower order term in S, i.e., (in the case of Laplacian principal part) define $Su = -\Delta u + cu$ for some constant $c > 0$. This gives a better approximation of $Lu = -\Delta u + \eta u$ and, since S has constant coefficients, the auxiliary problems can be still be solved by the mentioned fast solvers. Theorem 1 remains true, since $Qu = (\eta - c)u$ is still compact; it may be no more a positive operator, but the only arising difference is that in Corollary 1 we replace $\lambda_j(Q_S)$ by $|\lambda_j(Q_S)|$.

Remark 2 The relevance of the extension of the results of [10] on $\eta \in C(\overline{\Omega})$ to our more general case (15) is motivated, e.g., by the following model. Consider a reaction-diffusion equation

$$\begin{cases} -\Delta z + q(z) = f, \\ z|_{\partial\Omega} = 0, \end{cases} \tag{29}$$

where $q \in C^1(\mathbb{R})$ and there exists $2 < p < \frac{2d}{d-2}$ such that

$$0 \le q'(\xi) \le \alpha + \beta|\xi|^{p-2} \qquad (\forall \xi \in \mathbb{R}^d). \tag{30}$$

Here, q describes the rate of reaction, which is typically of polynomial order as required in (30). The restriction on p means that the continuous embedding $H_0^1(\Omega) \subset L^p(\Omega)$ holds, hence the above problem is well-posed in $H_0^1(\Omega)$. Then, the Newton linearization around some z_n leads to the linear problem of the form

$$\begin{cases} -\Delta u + \eta u = g \\ u|_{\partial\Omega} = 0 \end{cases} \tag{31}$$

where

$$\eta = q'(z_n) \in L^{p/(p-2)}(\Omega) \tag{32}$$

due to the above assumptions. That is, we obtain a problem of the type (14).

Remark 3 Owing to the equality $\|e_k\|_{\mathbf{A}_h} = \|A^{-1/2}r_k\|$, the estimate (18) implies a similar one for the residuals:

$$\left(\frac{\|r_k\|}{\|r_0\|}\right)^{\frac{1}{k}} \le \frac{C_1}{k^\alpha}$$

where $C_1 = C\,cond(A)$.

3.2 Elliptic systems

In this section, we prove that the previous results can be extended to certain systems of elliptic PDEs. For simplicity and also due to practical occurrence, we only include

Laplacian principal parts; however, the results remain similar when the principal parts have the form (19).

3.2.1 Symmetric systems

First let us consider systems of the form

$$\begin{cases} -\Delta u_i + \eta_{i1}u_1 + \dots \eta_{is}u_s = g_i, \\ u_i|_{\partial\Omega} = 0, \quad (i = 1, \dots, s), \end{cases} \tag{33}$$

where $\boldsymbol{H} = \{\eta_{ij}\}_{i,j=1}^{s}$ is a symmetric positive semidefinite variable coefficient matrix such that

$$\eta_{ij} \in L^{p/(p-2)}(\Omega) \qquad (\forall i, j \in \{1, \dots, s\}).$$

Such systems arise, e.g., in the Newton linearization of gradient systems: if a nonlinear reaction-diffusion system corresponds to a potential

$$\phi(u_1, ..., u_s) = \int_\Omega \Big(\frac{1}{2}\sum_{i=1}^{s} |\nabla u_i|^2 + F(u_1, ..., u_s)\Big),$$

then the linearized problems have the form (33), which extend (31)–(32) to systems and the gradient structure implies the symmetry of the Jacobians $\boldsymbol{H} = F'(u_1, ..., u_s)$.

We work with the space $L^p(\Omega)^s$ with the norm

$$\|u\|_{L^p(\Omega)^s} = \left(\sum_{j=1}^{s} \|u_j\|^2_{L^p(\Omega)}\right)^{1/2} \qquad (u = (u_1, \dots, u_s) \in L^p(\Omega)^s).$$

Let $H = L^2(\Omega)^s$; furthermore, $D := (H^1_{0,G})^s$, where $H^1_{0,G}$ was defined in subsection 3.1 before (19). Using notation $u = (u_1 \dots u_s)$, we define the operators

$$Su := \begin{pmatrix} -\Delta u_1 \\ . \\ . \\ . \\ -\Delta u_s \end{pmatrix} \quad (u \in D), \qquad Qu := \boldsymbol{H}u \quad (u \in H^1_0(\Omega)^s). \tag{34}$$

Clearly, S is a uniformly positive symmetric operator in H. In fact, from (20),

$$\langle Su, u\rangle \geq C_\Omega \sum_{i=1}^{s} \|u_i\|^2_{L^2(\Omega)} = C_\Omega \|u\|^2_H . \tag{35}$$

Then, the energy space H_S is well defined with

$$\langle u, v\rangle_S = \sum_{i=1}^{s} \int_\Omega \nabla u_i \cdot \nabla v_i, \qquad \|u\|_{H_S}^2 = \sum_{i=1}^{s} \int_\Omega |\nabla u_i|^2$$

and so $H_S = H_0^1(\Omega)^s$. Furthermore, by (22), we have that

$$\|u\|_{H_S}^2 \geq \frac{1}{\hat{c}^2} \sum_{i=1}^{s} \|u_i\|_{L^p(\Omega)}^2 = \frac{1}{\hat{c}^2} \|u\|_{L^p(\Omega)^s}^2. \tag{36}$$

Then, there exists a unique bounded linear operator $Q_S \colon H_0^1(\Omega)^s \to H_0^1(\Omega)^s$ such that

$$\langle Q_S u, v\rangle_S = \int_\Omega \sum_{i,j=1}^{s} \eta_{ij} u_j v_i. \tag{37}$$

It is easy to see that Q_S is self-adjoint in H_S. Analogously to (23), by (36), (35) and Hölder's inequality, we get

$$\begin{aligned}
\|Q_S v\|_{H_S} &= \sup_{\|u\|_S=1} |\langle Q_S v, u\rangle_S| \\
&\leq \sup_{\|u\|_{H_S}=1} \sum_{i,j=1}^{s} \int_\Omega |\eta_{ij}||v_j||u_i| \\
&\leq \sup_{\|u\|_{H_S}=1} \sum_{i,j=1}^{s} \|\eta_{ij}\|_{L^{p/(p-2)}(\Omega)} \|v_j\|_{L^p(\Omega)} \|u_i\|_{L^p(\Omega)} \\
&\leq M \sup_{\|u\|_{H_S}=1} \sum_{j=1}^{s} \|v_j\|_{L^p(\Omega)} \sum_{i=1}^{s} \|u_i\|_{L^p(\Omega)} \\
&\leq M \sup_{\|u\|_{H_S}=1} \sqrt{s} \left(\sum_{j=1}^{s} \|v_j\|_{L^p(\Omega)}^2 \right)^{1/2} \sqrt{s} \left(\sum_{i=1}^{s} \|u_i\|_{L^p(\Omega)}^2 \right)^{1/2} \\
&= Ms \sup_{\|u\|_{H_S}=1} \|v\|_{L^p(\Omega)^s} \|u\|_{L^p(\Omega)^s} \\
&\leq Ms\hat{c} \|v\|_{L^p(\Omega)^s},
\end{aligned} \tag{38}$$

where $M = \max_{i,j} \|\eta_{ij}\|_{L^{p/(p-2)}(\Omega)}$. Hence, we have proved that Q_S is a compact self-adjoint operator in H_S. Then, the characterization (24) of the eigenvalues of Q_S holds. The rest of the proof follows by modifying the scalar case. Now, instead of (25), we take the minimum in the following way over a smaller subset of finite rank operators:

$$\lambda_n(Q_S) \leq \min\{\|Q_S - Q_S L_{n-1}\| : \ L_{n-1} \in \mathcal{L}_{\text{diag}}(H_S), \ \text{rank}(L_{n-1}) \leq n-1\},$$

where we define $L_{n-1} \in \mathcal{L}_{\text{diag}}(H_S)$ by requiring

$$L_{n-1}u = \begin{pmatrix} L^s_{n-1}u_1 \\ . \\ . \\ . \\ L^s_{n-1}u_s \end{pmatrix}, \text{ such that } L^s_{n-1} \in \mathcal{L}(H^1_0(\Omega)) \text{ and } \operatorname{rank}(L^s_{n-1}) \le \left[\frac{n-1}{s}\right],$$

where [.] denotes the lower integer part. Furthermore, we shall use the approximation numbers

$$a_{\left[\frac{n-1}{s}\right]} = \min\left\{\|I - T_{n-1}\| : \ T_{n-1} \in \mathcal{L}(H^1_0(\Omega), L^p(\Omega)), \ \operatorname{rank}(T_{n-1}) \le \left[\frac{n-1}{s}\right]\right\}.$$

Note that if $n \le s$ then we can use the bound $\lambda_n(Q_S) \le \|Q_S\|$. For $n \ge s+1$, from (27), the above numbers are estimated by

$$a_{\left[\frac{n-1}{s}\right]} \le \hat{C}\left[\frac{n-1}{s}\right]^{-\alpha}, \tag{39}$$

with $\alpha = \frac{1}{d} - \frac{1}{2} + \frac{1}{p}$. Then,

$$\begin{aligned}
\|Q_S - Q_S L_{n-1}\| &= \sup_{u\in H_S} \frac{\|(Q_S - Q_S L_{n-1})u\|_{H_S}}{\|u\|_{H_S}} \\
&= \sup_{u\in H_S} \frac{\|Q_S(u - L_{n-1}u)\|_{H_S}}{\|u\|_{H_S}} \\
&\le Ms\hat{c} \sup_{u\in H_S} \frac{\|u - L_{n-1}u\|_{L^p(\Omega)^s}}{\|u\|_{H_S}} \\
&= Ms\hat{c} \sup_{u\in H_S} \frac{\left(\sum_{j=1}^s \|u_i - L^s_{n-1}u_i\|^2_{L^p(\Omega)}\right)^{1/2}}{\left(\sum_{j=1}^s \|u_i\|^2_{H^1_0(\Omega)}\right)^{1/2}} \\
&\le Ms\hat{c} \sup_{u\in H_S} \frac{\left(\|I - L^s_{n-1}\|^2_{\mathcal{L}(H^1_0(\Omega),L^p(\Omega))} \sum_{j=1}^s \|u_i\|^2_{H^1_0(\Omega)}\right)^{1/2}}{\left(\sum_{j=1}^s \|u_i\|^2_{H^1_0(\Omega)}\right)^{1/2}} \\
&= Ms\hat{c}\|I - L^s_{n-1}\|_{\mathcal{L}\left(H^1_0(\Omega),L^p(\Omega)\right)}.
\end{aligned}$$

Therefore,

$$\lambda_n(Q_S) \le Ms\hat{c}\min\left\{\|I - L^s_{n-1}\|_{\mathcal{L}\left(H^1_0(\Omega),L^p(\Omega)\right)} : \ L^s_{n-1} \in \mathcal{L}(H^1_0(\Omega), L^p(\Omega)),\ \operatorname{rank}(L^s_{n-1}) \le \left[\frac{n-1}{s}\right]\right\}$$
$$= Ms\hat{c}a_{\left[\frac{n-1}{s}\right]}.$$

Hence, by (39), we obtain the estimations

$$\lambda_n(Q_S) \le Ms\hat{c}\hat{C}\left[\frac{n-1}{s}\right]^{-\alpha} \qquad (\forall n \ge s+1), \tag{40}$$

$$\lambda_n(Q_S) \le \|Q_S\| \qquad (\forall n \le s). \tag{41}$$

Note that there exists $k_0, k_1 > 0$ such that

$$k_0 \le \frac{[x]}{x} \le k_1 \qquad (\forall x > 1)$$

(in fact, $k_0 = 1/2$ and $k_1 = 1$). Thus, for $n \ge s+1$,

$$\begin{aligned}\left[\frac{n-1}{s}\right]^{-\alpha} &\le \frac{1}{k_0^\alpha}\frac{s^\alpha}{(n-1)^\alpha}\\ &= \left(\frac{s}{k_0}\right)^\alpha\left(\frac{n^\alpha}{(n-1)^\alpha}\right)\frac{1}{n^\alpha}\\ &\le \left(\frac{(s+1)}{k_0}\right)^\alpha\frac{1}{n^\alpha}.\end{aligned}$$

Hence, (40) becomes

$$\lambda_n(Q_S) \le Ms\hat{c}\hat{C}\left(\frac{(s+1)}{k_0}\right)^\alpha\frac{1}{n^\alpha} := C_1\frac{1}{n^\alpha}$$

and by taking arithmetic means on both sides and splitting the sum, we get

$$\begin{aligned}\frac{1}{k}\sum_{n=1}^{k}\lambda_n(Q_S) &\le \frac{1}{k}\left(s\|Q_S\| + \sum_{n=s+1}^{k}\lambda_n(Q_S)\right)\\ &\le \frac{1}{k}\left(s\|Q_S\| + C_1\sum_{n=s+1}^{k}\frac{1}{n^\alpha}\right)\\ &\le \frac{1}{k}\left(s\|Q_S\| + C_1\int_s^k\frac{1}{x^\alpha}\,dx\right)\\ &\le \frac{s}{k}\|Q_S\| + \frac{C_1}{1-\alpha}\frac{1}{k^\alpha}\\ &\le C_2\frac{1}{k^\alpha},\end{aligned}$$

where $C_2 = \max\{s\|Q_S\|, C_1(1-\alpha)^{-1}\}$. Finally, by Corollary 1, we obtain that there exists $C > 0$ such that for all $k \in \mathbb{N}$

$$\left(\frac{\|e_k\|_{\mathbf{A}_h}}{\|e_0\|_{\mathbf{A}_h}}\right)^{\frac{1}{k}} \leq \frac{C}{k^{\alpha}} \tag{42}$$

with $\alpha = \frac{1}{d} - \frac{1}{2} + \frac{1}{p}$, that is, Theorem 1 holds in exactly the same from for the above systems of PDEs as well.

3.2.2 Extension to non-symmetric systems

Let us now study (33) for $\boldsymbol{H} = \{\eta_{i,j}\}_{i,j=1}^{s}$ non-symmetric. We apply the *generalized minimal residual (GMRES) method* to the corresponding discretized system. This method is the most widespread Krylov type iteration for non-symmetric systems, see, e.g., [13].

By [5], we have an analog of Corollary 1 when A is non-Hermitian. In this case the GMRES method provides superlinear convergence estimates for the residuals r_k, and (11) is replaced by the more general case (12). Altogether, we have

$$\left(\frac{\|r_k\|}{\|r_0\|}\right)^{1/k} \leq \frac{\|B^{-1}\|s}{k} \sum_{j=1}^{k} s_j(Q_S) \qquad (\forall k = 1, 2, \ldots, n). \tag{43}$$

To show that Theorem 1 still holds in this case, we follow the same steps as we did previously. We define the operators S, Q, Q_S as before in (34), (37). Here Q_S is no longer self-adjoint and its eigenvalues do not coincide with its singular values. Nonetheless, by [9, Ch.6, Th.1.5], we have the following characterization of the singular values of Q_S:

$$\forall n \in \mathbb{N}: \quad s_n(Q_S) = \min\{\|Q_S - L_{n-1}\| : \ L_{n-1} \in \mathcal{L}(H_S), \ \operatorname{rank}(L_{n-1}) \leq n-1\}. \tag{44}$$

Then, similarly to the proof for symmetric systems, we can see that there exists $C_1 > 0$ such that

$$\frac{1}{k}\sum_{n=1}^{k} s_n(Q_S) \leq \frac{C_1}{k^{\alpha}}, \qquad \text{where } \alpha = \frac{1}{d} - \frac{1}{2} + \frac{1}{p}. \tag{45}$$

Therefore, by (43), we obtain that there exists $C_2 > 0$ such that

$$\left(\frac{\|r_k\|}{\|r_0\|}\right)^{1/k} \leq \frac{C_2}{k^{\alpha}}. \tag{46}$$

3.2.3 The efficiency of the preconditioners

For elliptic systems, the auxiliary problem $\mathbf{S}w_k = \mathbf{Q}p_k$ is the FEM discretization of the elliptic system

$$\begin{cases} -\Delta(w_k)_1 &= \sum_{j=1}^{s} \eta_{1j}(p_k)_j, \\ -\Delta(w_k)_2 &= \sum_{j=1}^{s} \eta_{2j}(p_k)_j, \\ & \cdot \\ & \cdot \\ & \cdot \\ -\Delta(w_k)_s &= \sum_{j=1}^{s} \eta_{sj}(p_k)_j, \\ (w_i)|_{\partial\Omega} &= 0 \qquad (i = 1, \ldots, s) \end{cases}$$

where $w_k = ((w_k)_1, \ldots, (w_k)_s)$ is the unknown function and the right-hand side arises from the known functions $(p_k)_1, \ldots, (p_k)_s$. The main point is that (in contrast to the original one) this system is uncoupled, i.e., the above equations are independent of one another. Hence, they can be solved in parallel.

We note that the idea of Remark 1 can also be used here: one may include constant lower order terms in S, which is especially useful if $\boldsymbol{H}$ has large entries. Then, $-\Delta(w_k)_i$ above is replaced by $-\Delta(w_k)_i + c_i(w_k)_i$. For instance, we may set $c_i = 1/2\|\boldsymbol{H}\|$ or $c_i = 1/2\sum_{j=1}^{s} \eta_{ij}$.

In practice, these types of systems can be very large, e.g., in [16], long-range transport of air pollution models are described by a system of PDEs with $s = 30$. That is, whereas the original problem is a coupled PDE system of several components, the preconditioner leads to uncoupled problems corresponding to the FEM discretization of single PDEs, which is considerably cheaper. This shows the efficiency of the proposed preconditioners.

4 Numerical tests

Let us solve the following PDEs numerically:

$$\begin{cases} -\Delta u + \eta_1 u = f_i & \text{in } \Omega = [0, 1]^2, \\ u|_{\partial\Omega} = 0, \end{cases} \tag{47}$$

with $i = 1, 2$, and

$$\begin{cases} -\Delta u + \eta_2 u = f_1 & \text{in } \Omega = [0, 1]^2, \\ u|_{\partial\Omega} = 0, \end{cases} \tag{48}$$

where $p > 2$, and $\eta_1, \eta_2 \in L^{\frac{p}{p-2}}(\Omega)$ are defined as

$$\eta_1(x, y) = (x^2 + y^2)^{-\beta}$$

and

$$\eta_2(x, y) = ((x - 0.5)^2 + (y - 0.5)^2)^{-\beta}$$

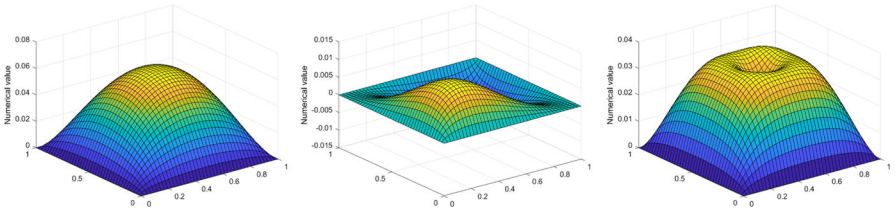

Fig. 1 Graphs of the numerical solutions with $N = 40$ for (47) with right-hand side f_i $(i = 1, 2)$ and $\beta = 1/4$, and for (48) with right-hand side f_1 and $\beta = 3/4$, respectively

for some $0 < \beta < \frac{p-2}{p}$. Furthermore,

$$f_1(x, y) = 1,$$

$$f_2(x, y) = 1 - x - y.$$

Applying the finite element method to (47) and (48) with stepsize $h = 1/(N + 1)$, we obtain the algebraic system

$$(\mathbf{G}_h + \mathbf{D}_h)\mathbf{c}_i = \mathbf{g}_h^i, \quad i = 1, 2, 3. \tag{49}$$

The cases $i = 1, 2$ and $i = 3$ refer to the FEM discretization of (47) and (48), respectively. Then, we apply $\mathbf{G}_h$ as a preconditioner and we solve the preconditioned system using the CGM. We used Courant elements and the computations were executed in Matlab (Fig. 1).

To measure the error of the PCGM, we use the energy norm

$$\|e\|_{\mathbf{A}_h} = \langle \mathbf{A}_h e, e\rangle^{\frac{1}{2}} \qquad (e \in \mathbb{R}^{N^2}),$$

Table 1 Norm of residual error r_k^i at each iteration of PCGM applied to system (49). Here $N = 40$ and $\beta = 1/4$

	$\|r_k^1\|_{\mathbf{G}_h}$	$\|r_k^2\|_{\mathbf{G}_h}$	$\|r_k^3\|_{\mathbf{G}_h}$
1	0.1872869890826060000000	0.043859195130465000000	0.187286989082606000000
2	0.002177821275260310000	0.000374462121567490000	0.005761180739756000000
3	0.000013427250794337400	0.000008998388683811820	0.000004194704792965480
4	0.000000122431712579675	0.000000061469010091297	0.000000150342994888464
5	0.000000000441761718591	0.000000000307598713847	0.000000000403838970113
6	0.000000000002105875799	0.000000000001126603119	0.000000000003834039672
7	0.000000000000008209336	0.000000000000003508771	0.000000000000119227495
8	0.000000000000000300119	0.000000000000000011073	0.000000000000001254764
9	0.000000000000000000081	0.000000000000000000086	0.000000000000000014252
10	0.000000000000000000006	0.000000000000000000005	0.000000000000000000098

Table 2 Values of $\hat{\delta}_k$ for different α's and β's, with a fixed mesh size. Here $N = 40$

	$\beta = 2/3, \alpha = 0.15$			$\beta = 3/4, \alpha = 0.12$			$\beta = 1/4, \alpha = 0.374$			$\beta = 1/2, \alpha = 0.24$		
	$\hat{\delta}_k^1$	$\hat{\delta}_k^2$	$\hat{\delta}_k^3$	$\hat{\delta}_k^1$	$\hat{\delta}_k^2$	$\hat{\delta}_k^3$	$\hat{\delta}_k^1$	$\hat{\delta}_k^2$	$\hat{\delta}_k^3$	$\hat{\delta}_k^1$	$\hat{\delta}_k^2$	$\hat{\delta}_k^3$
1	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
2	0.1786	0.1904	0.5319	0.1921	0.2098	0.6236	0.1397	0.1197	0.2273	0.1592	0.1593	0.3827
3	0.0925	0.1087	0.2623	0.1027	0.1127	0.3229	0.0627	0.0889	0.0905	0.0790	0.1026	0.1714
4	0.0621	0.0744	0.1509	0.0702	0.0792	0.1907	0.0478	0.0578	0.0503	0.0537	0.0670	0.0960
5	0.0476	0.0548	0.0973	0.0542	0.0611	0.1239	0.0344	0.0427	0.0337	0.0406	0.0476	0.0617
6	0.0372	0.0434	0.0751	0.0432	0.0490	0.0946	0.0293	0.0336	0.0323	0.0324	0.0376	0.0514
7	0.0313	0.0352	0.0941	0.0362	0.0406	0.1188	0.0256	0.0279	0.0375	0.0260	0.0304	0.0623
8	0.0264	0.0297	0.0754	0.0300	0.0340	0.0921	0.0231	0.0244	0.0368	0.0225	0.0254	0.0569
9	0.0227	0.0253	0.0749	0.0261	0.0287	0.0893	0.0207	0.0245	0.0368	0.0205	0.0225	0.0562
10	0.0203	0.0221	0.0743	0.0232	0.0250	0.0901	0.0213	0.0242	0.0352	0.0191	0.0208	0.0544

where $\mathbf{A}_h = \mathbf{G}_h + \mathbf{D}_h$. Table 1 shows the residual error obtained at each iteration $k \leq 10$ of the method applied to (49) for $i = 1, 2, 3$, respectively.

To test Theorem 1, note that $d = 2$ and so $\alpha = \frac{1}{p}$. Furthermore, recall that

$$\eta_1, \eta_2 \in L^{\frac{p}{p-2}}(\Omega) \quad \text{if } \beta < \frac{p-2}{p} = 1 - 2\alpha.$$

That is, if $p > \frac{2}{1-\beta}$, we get that the theorem holds when $\alpha < \frac{1-\beta}{2}$. Table 2 shows the values of

$$\hat{\delta}_k = \left(\frac{\|r_k\|_{G_h}}{\|r_0\|_{G_h}} \right)^{\frac{1}{k}} k^{\alpha}$$

Table 3 Values of $\hat{\delta}_k$ for different mesh sizes with $\beta = 3/4, \alpha = 0.12$

	$\hat{\delta}_k^1$			$\hat{\delta}_k^2$			$\hat{\delta}_k^3$		
	$N = 20$	$N = 40$	$N = 80$	$N = 20$	$N = 40$	$N = 80$	$N = 20$	$N = 40$	$N = 80$
1	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
2	0.1910	0.1921	0.1924	0.2080	0.2098	0.2103	0.5838	0.6236	0.6518
3	0.1013	0.1027	0.1031	0.1123	0.1127	0.1128	0.2865	0.3229	0.3499
4	0.0683	0.0702	0.0707	0.0785	0.0792	0.0794	0.1620	0.1907	0.2129
5	0.0519	0.0542	0.0549	0.0594	0.0611	0.0616	0.1037	0.1239	0.1412
6	0.0403	0.0432	0.0443	0.0466	0.0490	0.0499	0.1320	0.0946	0.0997
7	0.0333	0.0362	0.0373	0.0375	0.0406	0.0418	0.1069	0.1188	0.0999
8	0.0274	0.0300	0.0316	0.0310	0.0340	0.0353	0.1018	0.0921	0.1009
9	0.0234	0.0261	0.0279	0.0279	0.0287	0.0305	0.1001	0.0893	0.0802
10	0.0223	0.0232	0.0245	0.0245	0.0250	0.0268	0.1026	0.0901	0.0781

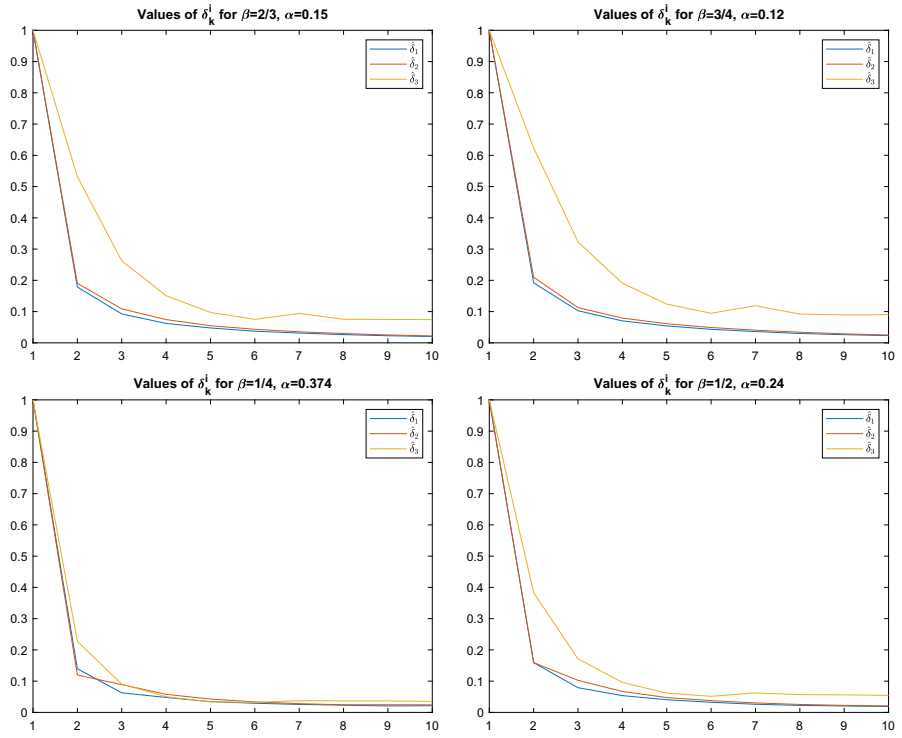

Fig. 2 Graphical representation of Table 2

for $i = 1, 2, 3$, respectively, with different choices of β and α while fixing a mesh size. The value of $\hat{\delta}_k^i$ $(i = 1, 2)$ corresponds to the system (47) with right-hand side f_i and the case $i = 3$ corresponds to the system (48). Note that residuals can be used when the exact solution is not known. In the symmetric case the bound (46) follows from the bound (18) owing to the equivalence of $\|e_k\|_{\mathbf{A}_h}$ and $\|r_k\|$, see Remark 3. Altogether, the estimate (46) is equivalent to requiring that $\hat{\delta}_k$ is bounded by some constant as k increases, and this is indeed demonstrated by Table 2 and Fig. 2.

Finally, Table 3 and Fig. 3 show the values of $\hat{\delta}_k$ for different mesh sizes while fixing the values of β. The numbers demonstrate that the results of Theorem 1 are not sensitive to the size of the mesh.

5 Summary and conclusions

We have studied the mesh independent superlinear convergence of preconditioned Krylov methods for the iterative solution of finite element discretizations of second-order elliptic boundary value problems. We have proved mesh independent estimations

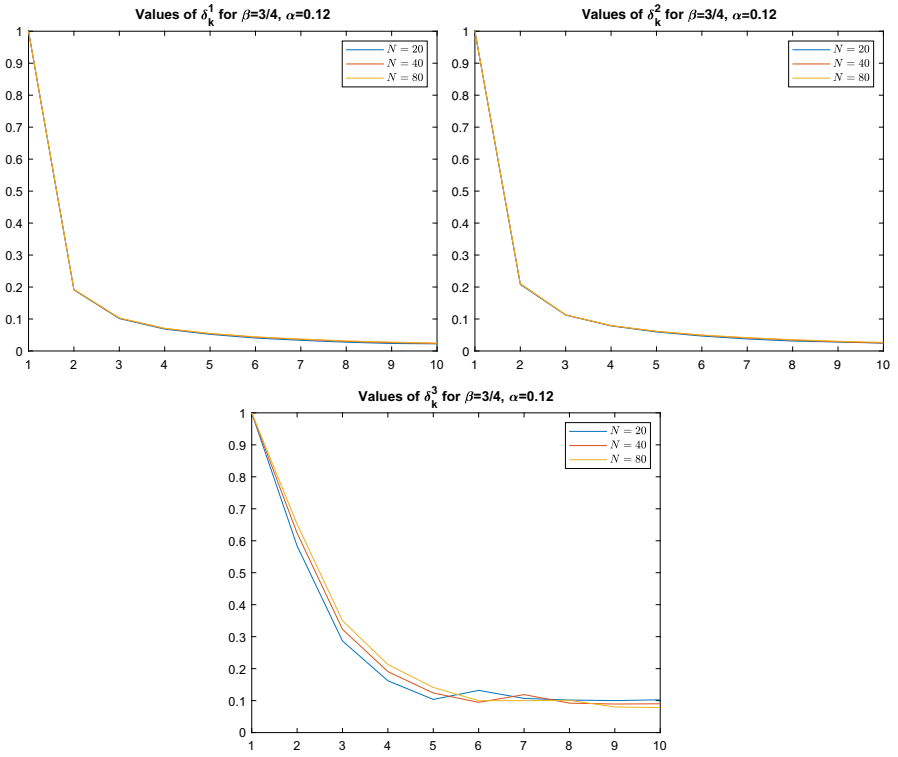

Fig. 3 Graphical representation of Table 3

for proper operator preconditioners for single equations and for systems, setting up a connection between the convergence rate and the Lebesgue exponent of the data. We have run numerical tests for equations with singular coefficients using different parameters. The tests have demonstrated the theoretical results.

Author contribution S.C. and J.K. derived the theoretical results and wrote the main manuscript text. S.C. wrote the codes.

Funding Open access funding provided by Eötvös Loránd University. This research has been supported by the Hungarian National Research, Development and Innovation Fund (NKFIH), under the funding scheme ELTE TKP 2021-NKTA-62 and the grants no. K137699 and SNN125119.

Data availability Not applicable

Declarations

Ethical approval Not applicable

Conflict of interest The authors declare no competing interests.

References

1. Axelsson, O.: Iterative solution methods. Cambridge University Press (1994)
2. Axelsson, O., Karátson, J.: Mesh independent superlinear PCG rates via compact-equivalent operators. SIAM J. Numer. Anal. **45**(4), 1495–1516 (2007)
3. Axelsson, O., Karátson, J.: Equivalent operator preconditioning for elliptic problems. Numer. Algorithms **50**(3), 297–380 (2009)
4. Axelsson, O., Karátson, J.: Superlinear convergence of the GMRES for pde-constrained optimization problems. Numer. Funct. Anal. Optim. **39**(9), 921–936 (2018)
5. Axelsson, O., Karátson, J., Magoules, F.: Robust superlinear Krylov convergence for complex non-coercive compact-equivalent operator preconditioners. SIAM J. Numer. Anal. **61**(2), 1057–1079 (2023)
6. Brézis, H.: Functional analysis, Sobolev spaces and partial differential equations, vol. 2. Springer (2011)
7. Chávez, G., Turkiyyah, G., Zampini, S., Ltaief, H., Keyes, D.: Accelerated cyclic reduction: a distributed-memory fast solver for structured linear systems. Parallel Comput. **74**, 65–83 (2018)
8. Edmunds, D.E., Triebel, H.: Entropy numbers and approximation numbers in function spacess. Proceedings of the London Mathematical Society **3**(1), 137–152 (1989)
9. I. Gohberg, S. Goldberg, M.A. Kaashoek: Operator theory: advances and applications. Classes of Linear Operators pp. 49 (1992)
10. Karátson, J.: Mesh independent superlinear convergence estimates of the conjugate gradient method for some equivalent self-adjoint operators. Appl Math **50**(3), 277–290 (2005)
11. Moret, I.: A note on the superlinear convergence of GMRES. SIAM J Numer. Anal. **34**(2), 513–516 (1997)
12. Rossi, T., Toivanen T.: A parallel fast direct solver for the discrete solution of separable elliptic equations. In: PPSC. Citeseer (1997)
13. Saad, Y., Schultz, M.H.: Gmres: a generalized minimal residual algorithm for solving nonsymmetric linear systems. SIAM J. Sci. Stat. Comp. **7**(3), 856–869 (1986)
14. Vybíral, J.: Widths of embeddings in function spaces. J. Complex. **24**(4), 545–570 (2008)
15. Winther, R.: Some superlinear convergence results for the conjugate gradient method. SIAM J. Numer. Anal. **17**(1), 14–17 (1980)
16. Zlatev, Z.: Numerical treatment of large air pollution models. In: Computer treatment of large air pollution models, pp. 69–109. Springer (1995)