Chapter 2
Efficient Resolution of Anisotropic Structures

Wolfgang Dahmen, Chunyan Huang, Gitta Kutyniok, Wang-Q Lim, Christoph Schwab, and Gerrit Welper

Abstract We highlight some results obtained in the DFG-SPP project "Adaptive Anisotropic Discretization Concepts". We focus on new developments concerning the *sparse representation* of possibly *high-dimensional* functions exhibiting strong *anisotropic* features and low regularity in isotropic Sobolev or Besov scales. Specifically, we focus on the solution of *transport equations* which exhibit *propagation of singularities* where, additionally, high-dimensionality enters when the convection field, and hence the solutions, depend on parameters varying over some compact set. Important constituents of our approach are directionally adaptive discretization concepts motivated by compactly supported shearlet systems, and well-conditioned stable variational formulations that support trial spaces with anisotropic refinements with arbitrary directionalities. We prove that they provide tight error-residual relations which are used to contrive rigorously founded adaptive refinement schemes which converge in L_2. Moreover, in the context of parameter dependent problems we discuss two approaches serving different purposes and working under different regularity assumptions. For "frequent query problems", making essential use of the novel well-conditioned variational formulations, a new *Reduced Basis Method* is outlined which exhibits a certain *rate-optimal*

W. Dahmen (✉)
RWTH Aachen, Templergraben 55, 52056 Aachen, Germany
e-mail: dahmen@igpm.rwth-aachen.de

C. Huang
School of Applied Mathematics, Central University of Finance and Economics,
39 South College Road, Haidian District Beijing, 100081 P.R. China
e-mail: hcy@cufe.edu.cn

G. Kutyniok • W.-Q. Lim
Technical University of Berlin, Straße des 17. Juni 136, 10623 Berlin, Germany
e-mail: kutyniok@math.tu-berlin.de; lim@math.tu-berlin.de

C. Schwab
ETH Zürich, HG G 57.1, Rämistr. 101, 8092 Zürich, Switzerland
e-mail: christoph.schwab@sam.math.ethz.ch

G. Welper
Texas A & M University, Mailstop 3368, College Station, TX 77843-3368, USA
e-mail: welper@math.tamu.edu

performance for indefinite, unsymmetric or singularly perturbed problems. For the radiative transfer problem with scattering a *sparse tensor* method is presented which mitigates or even overcomes the curse of dimensionality under suitable (so far still isotropic) regularity assumptions. Numerical examples for both methods illustrate the theoretical findings.

2.1 Introduction

The more complex a data site or mathematical model is the more adapted a corresponding mathematical representation needs to be in order to capture its information content at acceptable cost in terms of storage and computational complexity. In principle, this is true for mathematical objects described explicitly by large sets of possibly noisy or corrupted data but also for those given only implicitly as the solution of an operator equation. The latter scenario is perhaps even more challenging because direct observations are not possible. By "adapted representation" we mean a representation of the unknown function that exploits possibly global features of this function so as to require, for a prescribed target accuracy, only relatively few parameters to determine a corresponding approximation. Such global features could take a variety of forms such as (i) a high degree of regularity except at isolated singularities *located on lower dimensional manifolds*, or (ii) a particular *sparsity* possibly with respect to a *dictionary* which may even depend on the problem at hand. In fact, corresponding scenarios are not strictly disjoint. In either case reconstruction or approximation methods are necessarily nonlinear. For instance, as for (i), 1D *best N-term* wavelet approximations offer a powerful method based on selecting only possible few coefficients in an exact representation with respect to a given *universal background* dictionary, e.g. a wavelet basis. When dealing with more than one spatial variable the situation quickly becomes more complicated and for spatial dimensions much larger than three, classical numerical tools designed for the low dimensional regime become practically useless. This is commonly referred to as *curse of dimensionality*. Unfortunately, there seems to be no universal strategy of dealing with the curse of dimensionality, i.e., that works in all possible cases.

One global structural feature which is encountered in many multivariate scenarios is *anisotropy*: images, as functions of two variables, exhibit edges and discontinuities along curves. Higher dimensional biological images have sharp interfaces separating more homogeneous regions. Likewise highly anisotropic phenomena such as shear- or boundary layers are encountered in solutions to transport dominated initial-boundary value problems.

One major focus of the DFG-SPP project "Adaptive Anisotropic Discretization Concepts" has been to efficiently recover and economically encode *anisotropic* structures represented by explicitly given data or determined as solutions of operator equations which are prone to give rise to such structures. Regarding this latter case, which we will focus on in this article, *parametric transport problems* (as well as

close relatives) have served as guiding model problems for the following reasons:
(i) their solutions could exhibit shear or boundary layers and hence discontinuities
across lower dimensional manifolds calling for suitable *anisotropic discretizations*;
(ii) how to contrive suitable *variational formulations*, which in particular accom-
modate such anisotropic discretizations is much less clear than in the elliptic case;
(iii) *parametric versions* give rise to *high-dimensional* problems.

Concerning (i), *directional representation systems* like *curvelets* and *shearlets*
outperform classical *isotropic* wavelet bases when approximating so called "cartoon
images", see [23] and [9, 31–35]. For recent applications to imaging data, in
particular, *inpainting* as well as in combination with *geometric separation concepts*
the reader is referred to [24, 29]. In the present context of solving operator equations
we outline in Sect. 2.2 trial spaces which accommodate directional adaptivity. They
are motivated by recent constructions of compactly supported *piecewise polynomial
shearlet systems* (see e.g. [30]) because they are close to classical multiresolution
structures and similar in nature to classical discretization systems. Since cartoons
exhibit structural similarities with the solution to transport problems we state best
N-term error bounds for cartoon functions that will later serve as benchmarks for an
adaptive solver. For related anisotropic simplicial discretizations and their analysis
see e.g. [10, 12, 15].

As for (ii), our approach differs from previous works on anisotropic discretiza-
tions derived from "curvature information" on the current approximation and hence
not based on a rigorous error control (see e.g. [22] and the references therein), in
that we derive first in Sect. 2.3 *well conditioned variational formulations* for general
unsymmetric or indefinite and singularly perturbed problems, see [14, 16] for details
on *convection-diffusion* and *transport problems*. The underlying basic principles
are of independent interest by themselves and seem to have appeared first in [3].
They are also closely related to ongoing developments running under the flag of
Discontinuous Petrov Galerkin (DPG) Methods, see e.g. [19, 20]. The approach is
motivated by two crucial corner stones. On the one hand, one can essentially *choose*
the norm for the (infinite dimensional) trial space X by which one would like to
measure accuracy while adapting the norm for the (infinite dimensional) test space
Y so as to ensure that (ideally) the operator induced by this variational formulation
is even an isometry from X to Y' (the normed dual of Y). Numerical feasibility of
(nearly optimal) Petrov Galerkin discretizations based on such formulations, even
beyond a DPG framework, hinges on an appropriate *saddle point formulation* which
turns out to be actually crucial in connection with *model reduction* [18]. On the one
hand, this allows one to accommodate, for instance, L_2-frames. On the other hand,
the resulting tight error-residual relation is the basis of computable a-posteriori
error estimators [14, 16] and, ultimately, to rigorously founded adaptive anisotropic
refinement strategies.

These variational formulations apply in much more generality but in order to
address issue (iii) we exemplify them for the simple linear transport equation
(stationary or instationary) whose *parametric* version leads to high-dimensional
problems and forms a core constituent of kinetic models such as *radiative transport*.
There the transport direction – the parameter – varies over a unit sphere so that

solutions are functions of the spatial variables (and, possibly, of time) and of the transport direction.

We briefly highlight two ways of treating such parametric problems under slightly different objectives. Both strategies aim at approximating the solution $u(x, s)$, $x \in \Omega \subset \mathbb{R}^d$, $s \in S^{d-1}$, in the form

$$u(x, s) \approx \sum_{j=1}^{n} c_j(s) u_j(x). \tag{2.1}$$

In Sect. 2.4 the u_j are constructed *offline* in a greedy manner from *snapshots* of the solution manifold, thus forming a *solution dependent* dictionary. According to the paradigm of the *Reduced Basis Method* (RBM) the parameter dependent coefficients $c_j(s)$ are not given explicitly but can be efficiently computed in an *online* fashion, e.g. in the context of design or (online) optimization. This approach works the better the smoother the dependence of the solution on the parameters is so that the Kolmogorov n-widths decay rapidly with increasing n. Making essential use of the well conditioned variational formulations from Sect. 2.3, it can be shown that the resulting RBM has stability constants as close to one as one wishes yielding for the first time an RBM for transport and convection-diffusion problems with this property exhibiting the same rates as the Kolmogorov widths [18].

In Sect. 2.5 of this report, and in [26], we present algorithms which construct *explicitly* separable approximations of the form (2.1) for the parametric transport problem of radiative transfer. We also mention that separable approximations such as (2.1) arise in a host of other applications; for example, in parametric representations of PDEs with random field input data with the aid of sparse tensor product interpolation methods; we refer to [11, 13] and to the references therein. Adaptive near-minimal rank tensor solvers for problems in high dimensional phase space are established and analyzed in [2].

2.2 Anisotropic Approximations

Let $D = (0, 1)^2$ and let $\text{curv}(\partial\Omega)$ denote the curvature of $\partial\Omega \cap D$. The class of *cartoon-like functions* on $D = (0, 1)^2$,

$$\mathscr{C}(\zeta, L, M, D) := \{f_1 \chi_\Omega + f_2 \chi_{D \setminus \Omega} : \ \Omega \subset D, \ |\partial\Omega \cap D| \leq L, \ \partial\Omega \cap D \in C^2,$$

$$\text{curv}(\partial\Omega) \leq \zeta, \|f_i^{(l)}\|_{L_\infty(D)} \leq M, \ l \leq 2, \ i = 1, 2\}, \tag{2.2}$$

(where the parameters ζ, L are not mutually independent) has become a well accepted benchmark for sparse approximation in imaging [23]. Compactly supported shearlet systems for $L^2(\mathbb{R}^2)$ have been introduced in [30, 33] to provide
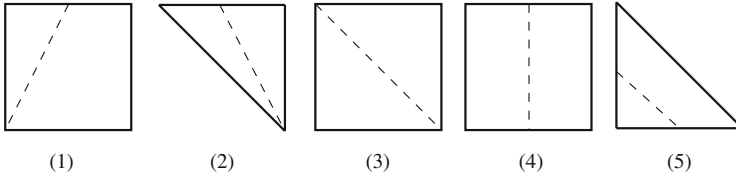
Fig. 2.1 Illustration of the partition rules

(near-) optimal sparse approximations for such classes. We observe that such cartoons also exhibit similar features as solutions to transport problems.

Unfortunately, even compactly supported shearlets do not comply well with quadrature and boundary adaptation tasks faced in variational methods for PDEs. We are therefore interested in generating locally refinable anisotropic partitions for which corresponding piecewise polynomial approximations realize the favorable near-optimal approximation rates for cartoon functions achieved by shearlet systems. Unfortunately, as shown in [36, Chapter 9.3], simple triangular bisections connecting the midpoint of an edge to the opposite vertex is not sufficient for warranting such rates, see [10, 15] for related work. In fact, a key feature would be to realize a "parabolic scaling law" similar to the shearlet setting. By this we mean a sufficient rapid directional resolution by anisotropic cells whose width scales like the square of the diameter. To achieve this we consider partitions comprised of triangles *and* quadrilaterals pointed out to us in Cohen and Mirebeau (2013, private communication). We sketch the main ideas and refer to [17] for details.

Starting from some initial partition consisting of triangles and quadrilaterals, *refined partitions* are obtained by splitting a given cell Q of a current partition according to one of the following rules:

  (i) Connect a vertex with the midpoint of an edge not containing the vertex.
 (ii) Connect two vertices.
(iii) Connect the midpoints of two edges which, when Q is a quadrilateral, do not share any vertex.

The types of bisections are indicated in Fig. 2.1: (1), (2) are examples of (i), (3) illustrates (ii), and (4), (5) are examples for (iii). One easily checks that these refinement rules produce only triangles and quadrilaterals. Moreover, a quadrilateral can be bisected in eight possible ways whereas a triangle can be split in six possible ways. Assigning to each split type a number in $I_Q = \{1, \ldots, 8\}$ when Q is a quadrilateral and a number in $I_Q = \{9, \ldots, 14\}$ when Q is a triangle, we denote by

$$R_{\iota_Q}(Q) = \{Q_1, Q_2\} \quad \text{for some } \ \iota_Q \in I_Q, \tag{2.3}$$

the refinement operator which replaces the cell Q by its two children Q_1, Q_2 generated, according to the choice ι_Q, by the above split rules (i)–(iii).

For any partition $\mathscr{G}$ of D, let $\mathbb{P}_1(\mathscr{G}) = \{v \in L_2(D) : v|Q \in \mathbb{P}_1, Q \in \mathscr{G}\}$ be the space of piecewise affine functions on $\mathscr{G}$ and denote by $\mathfrak{G}$ the set of all finite partitions that can be created by successive applications of R_{ι_Q} to define then

$$\Sigma_N := \bigcup \{\mathbb{P}_1(\mathscr{G}) : \mathscr{G} \in \mathfrak{G}, \#(\mathscr{G}) \leq N\}.$$

The next result from [17] shows that approximations by elements of Σ_N realize (and even slightly improve on) the known rates obtained for shearlet systems for the class of cartoon-like functions [33].

Theorem 2.1 ([17]). *Let $f \in \mathscr{C}(\zeta, L, M, D)$ with $D = (0, 1)^2$ and assume that the discontinuity curve $\Gamma = \partial\Omega \cap D$ is the graph of a C^2-function. Then one has*

$$\inf_{\varphi \in \Sigma_N} \|f - \varphi\|_{L_2(D)} \leq C(\zeta, L)M \, N^{-1} \log N,$$

where $C(\zeta, L)$ is an absolute constant depending only on ζ, L.

The proof of Theorem 2.1 is based on constructing a specific sequence $\mathscr{C}_j$ of admissible partitions from $\mathfrak{G}$ where the refinement decisions represented by R_{ι_Q} use full knowledge of the approximated function f. A similar sequence of partitions is employed in Sect. 2.3.4.2 where $\iota_Q \in I_Q$, however, results from an a posteriori criterion described later below. We close this section by a few remarks on the structure of the $\mathscr{C}_j$. Given $\mathscr{C}_{j-1}$, we first generate

$$\tilde{\mathscr{C}}_j = \{Q' \in \tilde{R}(Q) : Q \in \mathscr{C}_{j-1}\}, \tag{2.4}$$

where $\tilde{R}$ is either R_{ι_Q} or the identity. To avoid unnecessary refinements we define then $\mathscr{C}_j$ by replacing any pair of triangles $Q, Q' \in \tilde{\mathscr{C}}_j$ whose union forms a parallelogram P by P itself. This reduces the number of triangles in favor of parallelograms.

2.3 Well-Conditioned Stable Variational Formulations

In this section we highlight some new conceptual developments from [14, 16, 18] which are, in particular, relevant for the high dimensional parametric problems addressed later below.

2.3.1 The General Principles

Anisotropic structures are already exhibited by solutions of elliptic boundary value problems on polyhedral domains in 3D. However, related singularities are known a priori and can be dealt with by anisotropic *preset* mesh refinements. Anisotropic

structures of solutions to transport dominated problems can be less predictable so that a quest for *adaptive anisotropic* discretization principles gains more weight. Recall that every known rigorously founded adaptation strategy hinges in one way or the other on being able to relate a current error of an approximate solution to the corresponding *residual* in a suitable norm. While classical variational formulations of elliptic problems grant exactly such an error-residual relation, this is unclear for transport dominated problems. The first fundamental issue is therefore to *find* also for such problems suitable variational formulations yielding a well conditioned error-residual relation.

2.3.1.1 Abstract Petrov-Galerkin Formulation

Suppose that for a pair of Hilbert spaces X, Y (with scalar products $(\cdot, \cdot)_X, (\cdot, \cdot)_Y$ and norms $\|\cdot\|_X, \|\cdot\|_Y$), and a given bilinear form $b(\cdot, \cdot) : X \times Y \to \mathbb{R}$, the problem

$$b(u, v) = f(v), \quad v \in Y, \tag{2.5}$$

has for any $f \in Y'$ (the normed dual of Y) a unique solution $u \in X$. It is well-known that this is equivalent to the existence of constants $0 < c_b \leq C_b < \infty$ such that

$$\sup_{w \in X} \sup_{v \in Y} \frac{b(w, v)}{\|w\|_X \|v\|_Y} \leq C_b, \quad \inf_{w \in X} \sup_{v \in Y} \frac{b(v, w)}{\|w\|_X \|v\|_Y} \geq c_b, \tag{2.6}$$

and that, for each $v \in Y$, there exists a $w \in X$ such that $b(w, v) \neq 0$. This means that the operator $B : X \to Y'$, defined by $(Bu)(v) := b(u, v)$, $u \in X, v \in Y$, is an isomorphism with condition number $\kappa_{X,Y}(B) := \|B\|_{\mathscr{L}(X,Y')} \|B^{-1}\|_{\mathscr{L}(Y',X)} \leq C_b/c_b$. For instance, when (2.5) represents a *convection dominated convection-diffusion problem* with the classical choice $X = Y = H_0^1(\Omega)$, the quotient C_b/c_b becomes very large. Since

$$\|B\|_{\mathscr{L}(X,Y')}^{-1} \|Bv - f\|_{Y'} \leq \|u - v\|_X \leq \|B^{-1}\|_{\mathscr{L}(Y',X)} \|Bv - f\|_{Y'}, \tag{2.7}$$

the error $\|u - v\|_X$ can then *not* be tightly estimated by the residual $\|Bv - f\|_{Y'}$.

2.3.1.2 Renormation

On an abstract level the following principle has surfaced in a number of different contexts such as *least squares methods* (see e.g. [5, 8]) and the so-called, more recently emerged *Discontinuous Petrov Galerkin (DPG) methods*, see e.g. [3, 16, 19, 20] and the references therein. The idea is to fix a norm, $\|\cdot\|_Y$, say, and modify

the norm for X so that the corresponding operator even becomes an *isometry*. More precisely, define

$$\|u\|_{\hat{X}} := \sup_{v \in Y} \frac{b(u, v)}{\|v\|_Y} = \|Bu\|_{Y'} = \|R_Y^{-1} Bu\|_Y, \tag{2.8}$$

where $R_Y : Y \to Y'$ is the Riesz map defined by $(v, z)_Y = (R_Y v)(z)$. The following fact is readily verified, see e.g. [16, 40].

Remark 2.1. One has $\kappa_{\hat{X}, Y}(B) = 1$, i.e., (2.6) holds with $c_b = C_b = 1$ when $\| \cdot \|_X$ is replaced by $\| \cdot \|_{\hat{X}}$.

Alternatively, fixing X and redefining $\| \cdot \|_Y$ by $\|v\|_{\hat{Y}} := \|B^* v\|_{X'}$, one has $\kappa_{X, \hat{Y}}(B) = 1$, see [16]. Both possibilities lead to the error residual relations

$$\|u - w\|_X = \|f - Bw\|_{\hat{Y}'}, \quad \|u - w\|_{\hat{X}} = \|f - Bw\|_{Y'}, \quad u, w \in X. \tag{2.9}$$

2.3.2 Transport Equations

Several variants of these principles are applied and analyzed in detail in [14] for convection-diffusion equations. We concentrate in what follows on the limit case for vanishing viscosity, namely pure transport equations. For simplicity we consider the domain $D = (0, 1)^d, d = 1, 2, 3$, with $\Gamma := \partial D$, denoting as usual by $\boldsymbol{n} = \boldsymbol{n}(x)$ the unit outward normal at $x \in \Gamma$ (excluding the four corners, of course). Moreover, we consider velocity fields $\boldsymbol{b}(x), x \in D$, which for simplicity will always be assumed to be differentiable, i.e., $\boldsymbol{b}(x) \in C^1(\overline{D})^d$. Likewise $c(x) \in C^0(\overline{D})$ will serve as the reaction term in the *first order transport equation*

$$\boldsymbol{b} \cdot \nabla u + cu = f_\circ \text{ in } D, \quad u = g \text{ on } \Gamma_-, \tag{2.10}$$

where $\Gamma_\pm := \{x \in \partial D : \pm \boldsymbol{b}(x) \cdot \boldsymbol{n}(x) > 0\}$ denotes the *inflow, outflow boundary*, respectively. Furthermore, to simplify the exposition we shall always assume that $2c - \nabla \cdot \boldsymbol{b} \geq c_0 > 0$ in D holds.

A priori there does not seem to be any "natural" variational formulation. Nevertheless, the above principle can be invoked as follows. Following e.g. [16], one can show that the associated bilinear form with derivatives on the test functions

$$b(w, v) := \int_D w(-\boldsymbol{b} \cdot \nabla v + v(c - \nabla \cdot \boldsymbol{b})) \, dx, \tag{2.11}$$

is trivially bounded on $L_2(D) \times W_0(-\boldsymbol{b}, D)$, where

$$W_0(\mp \boldsymbol{b}, D) := \text{clos}_{\| \cdot \|_{W(b,D)}} \{v \in C^1(D) \cap C(\overline{D}), v |_{\Gamma_\pm} \equiv 0\}, \tag{2.12}$$

and

$$\|v\|_{W(\boldsymbol{b},D)} := \left(\|v\|_{L_2(D)}^2 + \int_D |\boldsymbol{b} \cdot \nabla v|^2 \, dx \right)^{1/2}. \tag{2.13}$$

Moreover, the trace $\gamma_-(v)$ exists for $v \in W_0(\boldsymbol{b}, D)$ in $L_2(\Gamma_-, |\boldsymbol{b} \cdot \boldsymbol{n}|)$, endowed with the norm $\|g\|_{L_2(\Gamma_\pm, |\boldsymbol{b} \cdot \boldsymbol{n}|)}^2 = \int_{\Gamma_\pm} |g|^2 |\boldsymbol{b} \cdot \boldsymbol{n}| ds$ so that

$$f(v) := (f_\circ, v) + \int_{\Gamma_-} g\gamma_-(v)|\boldsymbol{b} \cdot \boldsymbol{n}| ds \tag{2.14}$$

belongs to $(W_0(\boldsymbol{b}, D))'$ and the variational problem

$$b(u, v) = f(v), \quad v \in W_0(-\boldsymbol{b}, D) \tag{2.15}$$

possesses a unique solution in $L_2(D)$ which, when regular enough, coincides with the classical solution of (2.10), see [16, Theorem 2.2].

Moreover, since $X = L_2(D) = X'$, the quantity $\|v\|_Y := \|B^*v\|_{L_2(D)}$ is an equivalent norm on $W_0(-\boldsymbol{b}, D)$, see [16], and Remark 2.1 applies, i.e.,

$$\|B\|_{\mathcal{L}(L_2(D),(W_0(\boldsymbol{b},D))')} = \|B^*\|_{\mathcal{L}(W_0(\boldsymbol{b},D),L_2(D))} = 1, \tag{2.16}$$

see [16, Proposition 4.1]. One could also reverse the roles of test and trial space (with the inflow boundary conditions being then essential ones) but the present formulation imposes least regularity on the solution which will be essential in the next section. Note that whenever a PDE is written as a first order system, X can always be arranged as an L_2-space.

Our particular interest concerns the *parametric case*, i.e., the constant convection field $\boldsymbol{s}$ in

$$\begin{aligned} \boldsymbol{s} \cdot \nabla u(x, \boldsymbol{s}) + \kappa(x)u(x, \boldsymbol{s}) &= f_\circ(x), \quad x \in D \subset \mathbb{R}^d, \ d = 2, 3, \\ u(x, \boldsymbol{s}) &= g(x, \boldsymbol{s}), \ x \in \Gamma_-(\boldsymbol{s}), \end{aligned} \tag{2.17}$$

may *vary* over a set of directions $\mathscr{S}$ so that now the solution u depends also on the transport direction $\boldsymbol{s}$. In (2.17) and the following we assume that $\operatorname{ess\,inf}_{x \in D} \kappa(x) \geq 0$. Thus, for instance, when $\mathscr{S} = S^2$, the unit $2-$sphere, u is considered as a function of five variables, namely $d = 3$ spatial variables and parameters from a two-dimensional set $\mathscr{S}$. This is the simplest example of a kinetic equation forming a core constituent in *radiative transfer* models. The in- and outflow boundaries now depend on $\boldsymbol{s}$:

$$\Gamma_\pm(\boldsymbol{s}) := \{x \in \partial D : \mp \boldsymbol{s} \cdot \mathbf{n}(x) < 0\}, \qquad \boldsymbol{s} \in \mathscr{S}. \tag{2.18}$$

Along similar lines one can determine u as a function of x and s in $X = L_2(D \times \mathscr{S})$ as the solution of a variational problem with test space $Y := \mathrm{clos}_{\|\cdot\|_{W(D \times \mathscr{S})}}\{v \in C(\mathscr{S}, C^1(D)) : v|_{\Gamma_\pm} \equiv 0\}$ with $\|v\|^2_{W(D \times \mathscr{S})} := \|v\|^2_{L_2(D \times \mathscr{S})} + \int_{\mathscr{S} \times D} |s \cdot \nabla v|^2 dx ds$. Again this formulation requires minimum regularity. Since later we shall discuss yet another formulation, imposing stronger regularity conditions, we refer to [16] for details.

2.3.3 δ-Proximality and Mixed Formulations

It is initially not clear how to exploit (2.9) numerically since the perfect inf-sup stability on the infinite dimensional level is *not* automatically inherited by a given pair $X_h \subset X, Y_h \subset Y$ of equal dimension. However, given $X_h \subset \hat{X}$, one can identify the "ideal" test space $Y(X_h) = R_Y^{-1}B(X_h)$ which may be termed ideal because

$$\sup_{w \in X_h} \sup_{v \in Y(X_h)} \frac{b(w, v)}{\|w\|_X \|v\|_Y} = \inf_{w \in X_h} \sup_{v \in Y(X_h)} \frac{b(v, w)}{\|w\|_X \|v\|_Y} = 1, \qquad (2.19)$$

see [16]. In particular, this means that the solution $u_h \in X_h$ of the corresponding *Petrov-Galerkin* scheme

$$b(u_h, v) = f(v), \quad v \in Y(X_h), \qquad (2.20)$$

realizes the best $\hat{X}$-approximation to the solution u of (2.5), i.e.,

$$\|u - u_h\|_{\hat{X}} = \inf_{w \in X_h} \|u - w\|_{\hat{X}}. \qquad (2.21)$$

Of course, unless Y is an L_2 space, the ideal test space $Y(X_h)$ is, in general, not computable exactly. To retain stability it is natural to look for a numerically computable test space Y_h that is sufficiently close to $Y(X_h)$.

One can pursue several different strategies to obtain numerically feasible test spaces Y_h. When (2.5) is a discontinuous Galerkin formulation one can choose Y as a product space over the given partition, again with norms induced by the graph norm for the adjoint B^* so that the approximate inversion of the Riesz map R_Y can be *localized* [19, 20]. An alternative, suggested in [14, 16], is based on noting that by (2.8) the ideal Petrov Galerkin solution u_h from (2.20) is a *minimum residual* solution in Y', i.e., $u_h = \mathrm{argmin}_{w \in X_h} \|f - Bw\|_{Y'}$ whose normal equations read $(f - Bu_h, Bw)_{Y'} = 0, w \in X_h$. Since the inner product $(\cdot, \cdot)_{Y'}$ is numerically hard to access, one can write $(f - Bu_h, Bw)_{Y'} = \langle R_Y^{-1}(f - Bu_h), Bw \rangle$, where the dual pairing $\langle \cdot, \cdot \rangle$ is now induced by the standard L_2-inner product. Introducing as an auxiliary variable the "lifted residual"

$$y = R_Y^{-1}(f - Bu_h), \qquad (2.22)$$

or equivalently $(R_Y y)(v) = \langle R_Y y, v \rangle = (y, v)_Y = \langle f - Bu_h, v \rangle$, $v \in Y$, one can show that (2.20) is equivalent to the saddle point problem

$$
\begin{aligned}
\langle R_Y y, v \rangle + b(u_h, v) &= \langle f, v \rangle, \ v \in Y, \\
b(w, y) \qquad\qquad &= \quad 0, \quad w \in X_h,
\end{aligned}
\tag{2.23}
$$

which involves only standard L_2-inner products, see [16, 18].

Remark 2.2. When working with $X, \hat{Y}$ instead of $\hat{X}, Y$, one has $R_Y = BR_{\hat{X}}^{-1}B^*$ and hence, when $X = L_2(D)$ as in (2.11), one has $R_Y = BB^*$.

Since the test space Y is still infinite dimensional, a numerical realization would require finding a (possibly small) subspace $V \subset Y$ such that the analogous saddle point problem with Y replaced by V is still inf-sup stable. The relevant condition on V can be described by the notion of δ-*proximality* introduced in [16], see also [14]. We recall the formulation from [18]: $V \subset Y$ *is δ-proximal for* $X_h \subset \hat{X}$ *if, for some* $\delta \in (0, 1)$, with $P_{Y,V}$ denoting the Y-orthogonal projection from Y to V,

$$
\|(I - P_{Y,V})R_Y^{-1}Bw\|_Y \leq \delta \|R_Y^{-1}Bw\|_Y, \quad w \in X_h .
\tag{2.24}
$$

For a discussion of how to monitor or realize δ-proximality we refer to [14, 16], see also Sect. 2.3.5.

Theorem 2.2 ([14, 16, 18]). *Assume that for given $X_h \times V \subset X \times Y$ the test space V is δ-proximal for X_h, i.e. (2.24) is satisfied. Then, the solution $(u_{X_h,V}, y_{X_h,V}) \in X_h \times V$ of the saddle point problem*

$$
\begin{aligned}
\langle R_Y y_{X_h,V}, v \rangle + b(u_{X_h,V}, v) &= \langle f, v \rangle, \ v \in V, \\
b(w, y_{X_h,V}) \qquad\qquad &= \quad 0, \quad w \in X_h,
\end{aligned}
\tag{2.25}
$$

satisfies

$$
\|u - u_{X_h,V}\|_{\hat{X}} \leq \frac{1}{1 - \delta} \inf_{w \in X_h} \|u - w\|_{\hat{X}}.
\tag{2.26}
$$

and

$$
\|u - u_{X_h,V}\|_{\hat{X}} + \|y - y_{X_h,V}\|_Y \leq \frac{2}{1 - \delta} \inf_{w \in X_h} \|u - w\|_{\hat{X}}.
\tag{2.27}
$$

Moreover, one has

$$
\inf_{w \in X_h} \sup_{v \in V} \frac{b(w, v)}{\|v\|_Y \|q\|_{\hat{X}}} \geq \sqrt{1 - \delta^2}.
\tag{2.28}
$$

Finally, (2.25) *is equivalent to the Petrov-Galerkin scheme*

$$b(u_{X_h,V}, v) = f(v), \quad v \in Y_h := P_{Y,V}(R_Y^{-1} B(X_h)) = P_{Y,V}(Y(X_h)). \qquad (2.29)$$

The central message is that the Petrov-Galerkin scheme (2.29) can be realized *without* computing a basis for the test space Y_h, which for *each* basis function could require solving a problem of the size dim V, by solving instead the saddle point problem (2.25). Moreover, the stability of both problems is governed by the δ-proximality of V. As a by-product, in view of (2.22), the solution component $y_{X_h,V}$ approximates the exact lifted residual $R_Y^{-1}(f - Bu_{X_h,V})$ and, as pointed out below, can be used for an a posteriori error control.

The problem (2.25), in turn, can be solved with the aid of an *Uzawa iteration* whose efficiency relies again on δ-proximality. For $k = 0, \ldots,$ solve

$$\begin{aligned}
\langle R_Y y^k, v \rangle &= \langle f - Bu^k, v \rangle, \quad v \in V, \\
(u^{k+1}, w)_{\hat{X}} &= (u^k, w)_{\hat{X}} + \langle B^* y^k, w \rangle, \quad w \in X_h.
\end{aligned} \qquad (2.30)$$

Thus, each iteration requires solving a symmetric positive definite Galerkin problem in V for the approximate lifted residual.

Theorem 2.3 ([16, Theorem 4.3]). *Assume that* (2.24) *is satisfied. Then the iterates generated by the scheme* (2.30) *converge to* $u_{X_h,V}$ *and*

$$\|u_{X_h,V} - u^{k+1}\|_{\hat{X}} \le \delta \|u_{X_h,V} - u^k\|_{\hat{X}}, \quad k = 0, 1, 2, \ldots. \qquad (2.31)$$

2.3.4 Adaptive Petrov-Galerkin Solvers on Anisotropic Approximation Spaces

The benefit of the above saddle point formulation is not only that it saves us the explicit calculation of the test basis functions but that it provides also an *error estimator* based on the *lifted residual* $y_h = y_h(u_{X_h,V}, f)$ defined by the first row of (2.25).

2.3.4.1 Abstract δ-Proximinal Iteration

In fact, it is shown in [16] that when $V_h \subset Y$ is even δ-proximal for $X_h + B^{-1}F_h$, with some finite dimensional subspace $F_h \subset Y'$, one has

$$(1 - \delta)\|f_h - Bw\|_{Y'} \le \|y_h(w, f_h)\|_Y \le \|f_h - Bw\|_{Y'}, \quad w \in X_h, \qquad (2.32)$$

where $f_h \in F_h$ is an approximation of $f \in Y'$. The space F_h controls which components of f are accounted for in the error estimator. The term $f - f_h$ is a *data oscillation* error as encountered in adaptive finite element methods. It follows that the current error of the Petrov-Galerkin approximation $u_{X_h,V}$ is controlled from below and above by the quantity $\|y_h\|_Y$. This can be used to formulate the *adaptive* Algorithm 1 that can be proven to give rise to a *fixed error reduction* per step. Its precise formulation can be found in [16, § 4.2]. It is shown in [16, Proposition 4.7] that each refinement step in Algorithm 1 below reduces the error by a fixed fraction. Hence it terminates after finitely many steps and outputs an approximate solution $\bar{u}$ satisfying $\|u - \bar{u}\|_{\hat{X}} \leq \varepsilon$.

Algorithm 1 Adaptive algorithm

1: Set target accuracy ε, initial guess $\bar{u} = 0$, initial error bound $e = \|f\|_{Y'}$, parameters
   $\rho, \eta, \alpha_1, \alpha_2 \in (0, 1)$, initial trial and δ-proximal test spaces X_h, V_h;
2: **while** $e > \varepsilon$ **do** solve (2.25) within accuracy $\alpha_1 \rho$ (e.g. by an Uzawa iteration with initial guess
   $\bar{u}$) to obtain an approximate solution pair $(\hat{y}, \hat{u}) \in V_h \times X_h$;
3:     enlarge X_h to $X_{h,+}$ in such a way that

$$\inf_{g \in X_{h,+}} \|B^*\hat{y} - g\|_{\hat{X}'} \leq \eta \|B^*\hat{y}\|_{\hat{X}'} \quad \text{and set} \quad r := \operatorname{argmin}_{g \in X_{h,+}} \|B^*\hat{y} - g\|_{\hat{X}'}; \quad (2.33)$$

4:     compute $X_{h'} \supset X_h$, $F_{h'} \supset F_h$, $f_h \in BX_{h'} + F_{h'}$ such that $\|f - f_h\|_{Y'} \leq \alpha_2 \rho e$;
5:     set $X_h + X_{h,+} + X_{h'} \rightarrow X_h$, $\rho e \rightarrow e$, and choose a δ-proximal subspace V_h for X_h;
6: set $\hat{u} + r_X \rightarrow \bar{u}$.
7: **end while**

2.3.4.2   Application to Transport Equations

We adhere to the setting described in Sect. 2.3.2, i.e., $X = \hat{X} = L_2(D)$, $\hat{Y} = Y = W_0(-\boldsymbol{b}, D)$, and $R_Y = BB^*$.

The trial spaces that we now denote by X_j to emphasize the nested construction below, are spanned by discontinuous piecewise linear functions on a mesh composed of cells from collections $\mathscr{C}_j$, i.e.,

$$X_j = \mathbb{P}_1(\mathscr{C}_j), \quad j \geq 0, \quad (2.34)$$

where the collections $\mathscr{C}_j$ are derived from collections $\tilde{\mathscr{C}}_j$ of the type (2.4) as described in Sect. 2.2.

Given X_j of the form (2.34), the test spaces V_j are defined by

$$V_j := \mathbb{P}_2(\mathscr{G}_j) \cap C(D) \quad \text{with} \quad \mathscr{G}_j := \{R^{iso}(Q) : Q \in \mathscr{C}_j\}, \quad (2.35)$$

where $R^{iso}(Q) = \{Q \cap P_i : i = 1, \ldots, 4\}$ is defined as follows. Let P be a parallelogram containing Q and sharing at least three vertices with Q. (There exist at most two such parallelograms and we choose one of them). Then the parallelograms P_i result from a dyadic refinement of P. As pointed out later, the

test spaces V_j constructed in this way, appear to be sufficiently large to ensure δ-proximity for X_j for δ significantly smaller than one uniformly with respect to j.

Since the test spaces V_j are determined by the trial spaces X_j the crucial step is to generate X_{j+1} by enlarging X_j based on an a posteriori criterion that "senses" directional information. This, in turn, is tantamount to a possibly anisotropic adaptive refinement of $\mathscr{C}_j$ leading to the updated spaces for the next iteration sweep of the form (2.30). The idea is to use a greedy strategy based on the largest "fluctuation coefficients". To describe this we denote for each $\iota_Q \in I_Q$ by $\Psi_{R_{\iota_Q}(Q)}$ an orthonormal Alpert-type wavelet basis for the difference space $\mathbb{P}_1(R_{\iota_Q}(Q)) \ominus \mathbb{P}_1(Q)$, see [1]. We then set

$$\Psi_j = \{\psi_\gamma \in \Psi_{R(Q)} : Q \in \mathscr{C}_{j-1}\}, \tag{2.36}$$

where $\Psi_{R(Q)} = \bigcup_{\iota_Q \in I_Q} \Psi_{R_{\iota_Q}(Q)}$. Initializing $\mathscr{C}_0$ as a uniform partition (on a low level), we define for some fixed $\theta \in (0, 1)$

$$T_j = \theta \cdot \max_{\psi_\gamma \in \Psi_j} |\langle B^* r_j^K, \psi_\gamma \rangle|$$

for $j > 0$, where Ψ_j is the two level basis defined in (2.36) and $r_j^K = y^K$ is the lifted residual from the first row of the Uzawa iteration. Then, for each $Q \in \mathscr{C}_{j-1}$, we define its refinement $\tilde{R}(Q)$ (see the remarks following (2.4)) by

$$\tilde{R}(Q) := \begin{cases} \{Q\}, & \text{if} \quad \max_{\psi_\gamma \in \Psi_{R(Q)}} |\langle B^* r_j^K, \psi_\gamma \rangle| \le T_j, \\ R_{\hat{\iota}_Q}(Q), & \text{otherwise}, \end{cases}$$

where $\hat{\iota}_Q$ is chosen to maximize $\max_{\psi_\gamma \in \Psi_{R_{\iota_Q}(Q)}} |\langle B^* r_j^K, \psi_\gamma \rangle|$ among all $\iota_Q \in I_Q$. One can then check whether this enrichment yields a sufficiently accurate L_2-approximation of $B^* r_j^K$ (step 3 of Algorithm 1). In this case, we adopt $\mathscr{C}_j$. Otherwise, the procedure is repeated for a smaller threshold θ.

2.3.5 Numerical Results

We provide some numerical experiments to illustrate the performance of the previously introduced anisotropic adaptive scheme for first order linear transport equations and refer to [17] for further tests. We monitor δ-proximity by computing

$$\frac{\inf_{\phi \in V_j} \|u_j - u_j^K - B^* \phi\|_{L_2([0,1]^2)}}{\|u_j - u_j^K\|_{L_2([0,1]^2)}}, \tag{2.37}$$
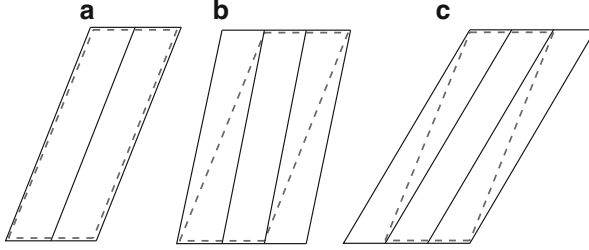
Fig. 2.2 Possible directional adjustments are illustrated for a parallelogram P (*dashed line*). (**a**) Rule (iii) of Sect. 2.2 yields two parallelograms with the same "direction". (**b**), (**c**) Applying rule (i) twice, changes the anisotropic direction slightly. The three refined parallelograms depicted in (**b**), (**c**) illustrate the results of a possible merging of adjacent triangles

where $u_j = \mathrm{argmin}_{v_j \in X_j} \|u - v_j\|_{L_2(D)}$. This is only a lower bound of the δ-proximality constant δ for one particular choice of w in (2.24) which coincides with the choice of w in the proof in [16]. In the following experiment, the number K of Uzawa iterations is for simplicity set to $K = 10$. One could as well employ an early termination of the inner iteration based on a posteriori control of the lifted residuals r_j^k.

We consider the transport equation (2.10) with zero boundary condition $g = 0$, convection field $\boldsymbol{b} = (x_2, 1)^T$, and right hand side $f = \chi_{\{x_1 > x_2^2/2\}} + 1/2 \cdot \chi_{\{x_1 \le x_2^2/2\}}$ so that the solution exhibits a discontinuity along the curvilinear shear layer given by $x_1 = \frac{1}{2}x_2^2$.

In this numerical example we actually explore ways of reducing the relatively large number of possible splits corresponding to the operators $R_{\iota_Q}, \iota_Q \in I_Q$, while still realizing the parabolic scaling law. In fact, we confined the cells to intersections of parallelograms P and their intersections with the domain D, much in the spirit of shearlet systems, employing anisotropic refinements as illustrated in Fig. 2.2 as well as the isotropic refinement R^{iso}. Permitting occasional overlaps of parallelograms, one can even avoid any interior triangles, apparently without degrading the accuracy of the adaptive approximation. The general refinement scheme described in Sect. 2.2 covers the presently proposed one as a special case, except, of course, for the possible overlap of cells.

Figure 2.3a, b show the adaptive grids associated with the trial space X_5 and the test space V_5. The refinement in the neighborhood of the discontinuity curve reflects a highly anisotropic structure. Figure 2.3c illustrates the approximation given by 306 basis elements. We emphasize that the solution is very smooth in the vicinity of the discontinuity curve and oscillations across the jump are almost completely absent and in fact much less pronounced than observed for isotropic discretizations. Figure 2.3d indicates the optimal rate realized by our scheme, see Theorem 2.1. The estimated values of the proximality parameter δ, displayed in Table 2.1, indicate the numerical stability of the scheme.

a

b

c

d
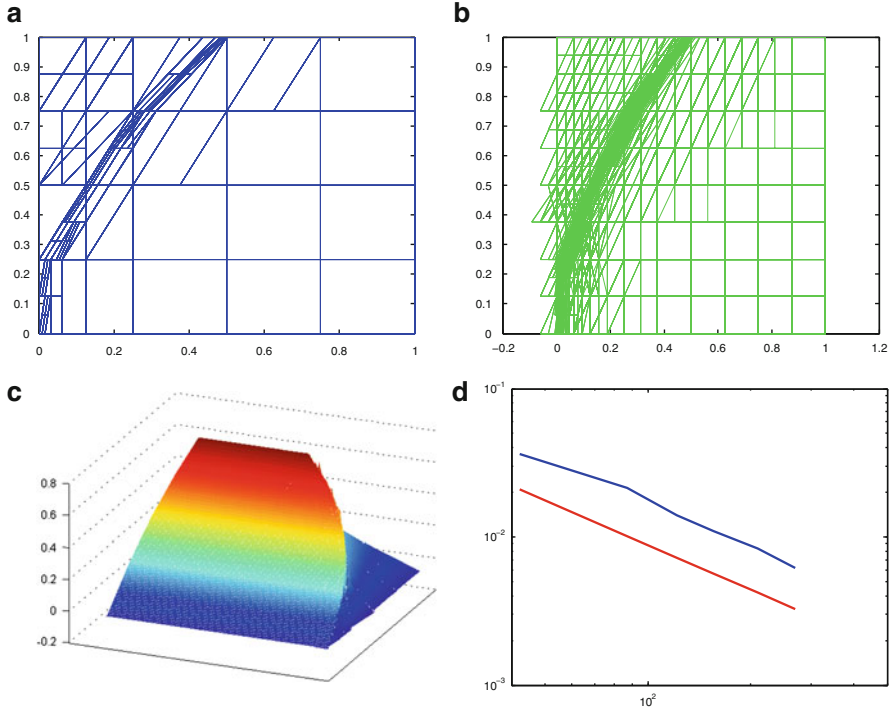
Fig. 2.3 (**a**) Adaptive grid for the trial space X_5. (**b**) Adaptive grid for the test space V_5. (**c**) Approximate solution (306 basis elements). (**d**) $L^2(D)$ errors (vertical axis) for N degrees of freedom (horizontal axis) achieved by the adaptive scheme (*blue*) in comparison with the optimal rate N^{-1} (*red*), predicted by Theorem 2.1. This is to be compared with the rate $N^{-1/2}$ realized by adaptive *isotropic* refinements [16]

Table 2.1 Numerical estimates (2.37) for the proximality constant δ and for the L_2 approximation error

n	Estimated δ	$\|u_j^K - u\|_{L_2([0,1]^2)}$
48	0.298138	0.036472
99	0.442948	0.021484
138	0.352767	0.013948
177	0.322156	0.010937
237	0.316545	0.008348
306	0.307965	0.006152

In the remainder of the paper we discuss parametric equations whose solutions are functions of spatial variables and additional parameters. Particular attention will here be paid to the *radiative transfer problems*, where the dimension of the physical domain is 2 or 3.

2.4   Reduced Basis Methods

2.4.1   Basic Concepts and Rate Optimality

Model reduction is often necessary when solutions to *parametric families* of PDEs
are frequently queried for different parameter values e.g. in an online design or
optimization process. The linear transport equation (2.17) is a simple example of
such a *parameter dependent* PDE. Since (a) propagation of singularities is present
and (b) the parameters determine the propagation direction s it turns out to already
pose serious difficulties for standard model reduction techniques.

   We emphasize that, rather than considering a *single* variational formulation for
functions of spatial variables and parameters, as will be done later in Sect. 2.5, we
take up the parametric nature of the problem by considering a *parametric family
of variational formulations*. That is, for each fixed s the problem is an ordinary
linear transport problem for which we can employ the corresponding variational
formulation from Sect. 2.3.2, where now the respective spaces may depend on the
parameters. In this section we summarize some of the results from [18] which are
based in an essential way on the concepts discussed in the previous section.

   In general, consider a family

$$b_\mu(u, v) = f(v), \quad u \in X_\mu, \, v \in Y_\mu, \, \mu \in \mathscr{P}, \, b_\mu(u, v) = \sum_{k=1}^{M} \Theta_k(\mu) b_k(u, v)$$

(2.38)

of well-posed problems, where $\mathscr{P} \subset \mathbb{R}^P$ is a compact set of parameters μ, and
the parameter dependence is assumed to be *affine* with smooth functions Θ_k. The
solutions $u(\cdot; \mu) = u(\mu)$ then become functions of the spatial variables and of the
parameters $\mu \in \mathscr{P}$.

   As before we can view (2.38) as a parametric family of operator equations
$B_\mu u = f$, where $B_\mu : X_\mu \to Y'_\mu$ is again given by $(B_\mu u)(v) = b_\mu(u, v)$. Each
particular solution $u(\mu)$ is a *point* on the *solution manifold*

$$\mathscr{M} := \{B_\mu^{-1} f : \mu \in \mathscr{P}\}.$$

(2.39)

Rather than viewing $u(\mu)$ as a point in a very high-dimensional (in fact infinite
dimensional) space, and calling a standard solver for each evaluation in a frequent
query problem, the *Reduced Basis Method* (RBM) tries to exploit the fact that each
$u(\mu)$ belongs to a much smaller dimensional manifold $\mathscr{M}$. Assuming that all the
spaces X_μ are equivalent to a reference Hilbert space X with norm $\| \cdot \|_X$, the
key objective of the RBM is to construct a possibly small dimensional linear space
$X_n \subset X$ such that for a given target accuracy $\varepsilon > 0$

$$\sup_{\mu \in \mathscr{P}} \inf_{w \in X_n} \|u(\mu) - w\|_X := \text{maxdist}_X(\mathscr{M}, X_n) \leq \varepsilon.$$

(2.40)

Once X_n has been found, bounded linear functionals of the exact solution $u(\mu)$ can be approximated within accuracy ε by the functional applied to an approximation from X_n, which, when n is small, can hopefully be determined at very low cost. The computational work in an RBM is therefore divided into an *offline* and an *online* stage. Finding X_n is the core offline task which is allowed to be computationally (very) expensive. More generally, solving problems in the "large" space X is part of the offline stage. Of course, solving a problem in X is already idealized. In practice X is replaced by a possibly very large trial space, typically a finite element space, which is referred to as the *truth* space and should be chosen large enough to guarantee the desired target accuracy, ideally certified by a posteriori bounds.

The computation of a (near-)best approximation $u_n(\mu) \in X_n$ is then to be *online feasible*. More precisely, one seeks to obtain a representation

$$u_n(\mu) = \sum_{j=1}^{n} c_j(\mu)\phi_j, \tag{2.41}$$

where the ϕ_j form a basis for X_n and where for each query $\mu \in \mathscr{P}$ the expansion coefficients $c_j(\mu)$ can be computed by solving only problems of the size n, see e.g. [39] for principles of practical realizations. Of course, such a concept pays off when the dimension $n = n(\varepsilon)$, needed to realize (2.40), grows very slowly when ε decreases. This means that the elements of $\mathscr{M}$ have sparse representations with respect to certain *problem dependent* dictionaries.

The by now most prominent strategy for constructing "good" spaces X_n can be sketched as follows. Evaluating for a given X_n the quantity $\text{maxdist}_X(\mathscr{M}, X_n)$ is infeasible because this would require to determine for *each* $\mu \in \mathscr{P}$ (or for *each* μ in a large training set $\mathscr{P}_h \subset \mathscr{P}$ which for simplicity we also denote by $\mathscr{P}$) the solution $u(\mu)$ which even for the offline stage is way too expensive. Therefore, one chooses a *surrogate* $R_n(\mu)$ such that

$$\inf_{w \in X_n} \|u(\mu) - w\|_X \le R_n(\mu, X_n), \quad \mu \in \mathscr{P}, \tag{2.42}$$

where the evaluation of $R_n(\mu, X_n)$ is fast and an optimization of $R_n(\mu, X_n)$ can therefore be performed in the offline stage. This leads to the *greedy algorithm* in Algorithm 2. A natural question is to ask how the spaces X_n constructed in such a greedy fashion compare with "best spaces" in the sense of the *Kolmogorov n-widths*

$$d_n(\mathscr{M})_X := \inf_{\dim W_n = n} \sup_{w \in \mathscr{M}} \inf_{z \in W_n} \|w - z\|_X . \tag{2.44}$$

The n-widths are expected to decay the faster the more regular the dependence of $u(\mu)$ is on μ. In this case an RBM has a chance to perform well.

Clearly, one always has $d_n(\mathscr{M})_X \le \text{maxdist}_X(\mathscr{M}, X_n)$. Unfortunately, the best constant C_n for which $\text{maxdist}_X(\mathscr{M}, X_n) \le C_n d_n(\mathscr{M})_X$ is $C_n = 2^n$, see [4, 6]. Nevertheless, when comparing *rates* rather than individual values, one arrives at

---

Algorithm 2 Greedy algorithm

1: **function GA**
2:     Set $X_0 := \{0\}$, $n = 0$,
3:     **while** $\text{argmax}_{\mu \in \mathscr{P}} R(\mu, X_n) \geq \varepsilon$ **do**
4:

$$
\begin{aligned}
\mu_{n+1} &:= \text{argmax}_{\mu \in \mathscr{P}} R(\mu, X_n), \\
u_{n+1} &:= u(\mu_{n+1}), \\
X_{n+1} &:= \text{span}\left\{X_n, \{u(\mu_{n+1})\}\right\} = \text{span}\{u_1, \ldots, u_{n+1}\}
\end{aligned}
\tag{2.43}
$$

5:     **end while**
6: **end function**

---

more positive results [4, 21]. The following consequence of these results asserts optimal performance of the greedy algorithm provided that the surrogate sandwiches the error of best approximation.

Theorem 2.4 ([18, Theorem 1.3]). *Assume that there exists a constant* $0 < c_R \leq 1$ *such that one has for all n*

$$
c_R R_n(\mu, X_n) \leq \inf_{w \in X_n} \|u(\mu) - w\|_X \leq R_n(\mu, X_n), \quad \mu \in \mathscr{P}. \tag{2.45}
$$

Then, the spaces X_n *produced by Algorithm 2 satisfy*

$$
d_n(\mathscr{M})_x \leq Cn^{-\alpha} \quad \implies \quad \text{maxdist}_X(\mathscr{M}, X_n) \leq \bar{C}n^{-\alpha}, \tag{2.46}
$$

where $\bar{C}$ *depends only on* C, α, *and* $\kappa(R_n) := 1/c_R$, *the* condition *of the surrogate.*

We call the RBM *rate-optimal* whenever (2.46) holds for any $\alpha > 0$. Hence, finding rate-optimal RBMs amounts to finding *feasible well-conditioned surrogates*.

2.4.2   A Double Greedy Method

Feasible surrogates that do not require the explicit computation of truth solutions for each $\mu \in \mathscr{P}$ need to be based in one way or the other on *residuals*. When (2.38) is a family of uniformly X-elliptic problems so that B_μ are uniformly bounded isomorphisms from X onto X', residuals indeed lead to feasible surrogates whose condition depends on the ratio of the continuity and coercivity constant. This follows from the mapping property of B_μ, stability of the Galerkin method, and the best approximation property of the Galerkin projection, see [18].

When the problems (2.38) are indefinite or unsymmetric and singularly perturbed these mechanisms no longer work in this way, which explains why the conventional RBMs do not perform well for transport dominated problems in that they are far from rate-optimal.

As shown in [18], a remedy is offered by the above *renormation principle* providing well-conditioned variational formulations for (2.38). In principle, these allow one to relate errors (in a norm of choice) to residuals in a suitably adapted dual norm which are therefore candidates for surrogates. The problem is that, given a trial space X_n, in particular a space generated in the context of an RBM, it is not clear how to obtain a sufficiently good test space such that the corresponding Petrov-Galerkin projection is comparable to the best approximation. The new scheme developed in [18] is of the following form:

(I)   Initialization: take $X_1 := \text{span}\{u(\mu_1)\}$, μ_1 randomly chosen, $Y_1 := \{0\}$;
(II) Given a pair of spaces $X_n, \tilde{V}_n$, the routine UPDATE-INF-SUP-δ enriches $\tilde{V}_n$ to a larger space V_n which is δ-proximal for X_n;
(III) Extend X_n to X_{n+1} by a greedy step according to Algorithm 2, set $\tilde{V}_{n+1} = V_n$, and go to (II) as long as a given target tolerance for an a posteriori threshold is not met.

The routine UPDATE-INF-SUP-δ works roughly as follows (see also [25] in the case of the Stokes system). First, we search for a parameter $\bar{\mu} \in \mathscr{P}$ and a function $\bar{w} \in X_n$ for which the inf-sup condition is worst, i.e.

$$\sup_{v \in \tilde{V}_n} \frac{b_{\bar{\mu}}(\bar{w}, v)}{\|v\|_{Y_{\bar{\mu}}} \|\bar{w}\|_{\hat{X}_{\bar{\mu}}}} = \inf_{\mu \in \mathscr{P}} \left(\inf_{w \in X_n} \sup_{v \in \tilde{V}_n} \frac{b_{\mu}(w, v)}{\|v\|_{Y_{\mu}} \|w\|_{\hat{X}_{\mu}}} \right). \qquad (2.47)$$

If this worst case inf-sup constant does not exceed yet a desired uniform lower bound, $\tilde{V}_n$ does not contain an effective *supremizer*, i.e., a function realizing the supremum in (2.47), for $\bar{\mu}, \bar{w}$, yet. However, since the truth space satisfies a uniform inf-sup condition, due to the same variational formulation, there exists a good supremizer in the truth space which, is given by the Galerkin problem

$$\bar{v} = R_{Y_{\bar{\mu}}}^{-1} B_{\bar{\mu}} \bar{w} = \text{argmax}_{v \in Y_{\bar{\mu}}} \frac{b_{\bar{\mu}}(\bar{w}, v)}{\|v\|_{Y_{\bar{\mu}}} \|\bar{w}\|_{\hat{X}_{\bar{\mu}}}},$$

providing the enrichment $\tilde{V}_n \to \text{span}\{\tilde{V}_n, R_{Y_{\mu}}^{-1} B_{\mu} \bar{w}\}$.

The *interior greedy stabilization loop* (II) ensures that the input pair X_n, Y_n in step (III) is inf-sup stable with an inf-sup constant as close to one as one wishes, depending on the choice of $\delta < 1$. By Theorem 2.2, each solution $u_n(\mu)$ of the discretized system for $(X_h, V) = (X_n, V_n)$ satisfies the near-best approximation property (2.26) and (2.27). Hence $\|f - B_{\mu}u_n(\mu)\|_{Y_{\mu}'}$ is a well conditioned surrogate (with condition close to one). Therefore, the assumptions of Theorem 2.4 hold so that the *outer greedy* step (III) yields a rate-optimal update. In summary, under the precise assumptions detailed in [18], the above *double greedy scheme* is *rate-optimal*.

Before turning to numerical examples, a few comments on the interior greedy loop UPDATE-INF-SUP-δ are in order.

(a) Finding $\bar{\mu}$ in (2.47) requires for each μ-query to perform a singular value decomposition in the low dimensional reduced spaces so that this is offline feasible, see [18, Remark 4.2].

(b) When the test spaces Y_μ all agree with a reference Hilbert space Y as sets and with equivalent norms it is easy to see that the interior stabilization loop terminates after at most M steps where M is the number of parametric components in (2.38), see [18, Remark 4.9] and [25, 38]. If, on the other hand, the spaces Y_μ differ even as sets, as in the case of transport equations when the transport direction is the parameter, this is not clear beforehand. By showing that the inf-sup condition is equivalent to a δ-proximality condition one can show under mild assumptions though that the greedy interior loop still terminates after a number of steps which is independent of the truth dimension, [18, Remark 4.11].

(c) In this latter case the efficient evaluation of $\| f - B_\mu u(\mu) \|_{Y'_\mu}$ requires additional efforts, referred to as *iterative tightening*, see [18, Section 5.1].

(d) The renormation strategy saves an expensive computation of stability constants as in conventional RBMs since, by construction, through the choice of δ, the stability constants can be driven as close to one as one wishes.

The scheme has been applied in [18] to convection-diffusion and pure transport problems where the convection directions are parameter dependent. Hence the variational formulations are of the form (2.38). We briefly report some results for the transport problem since this is an extreme case in the following sense. The test spaces Y_μ do *not* agree as sets when one would like the X_μ to be equivalent for different parameters. Hence, one faces the obstructions mentioned in (b), (c) above. Moreover, for discontinuous right hand side and discontinuous boundary conditions the dependence of the solutions on the parameters has low regularity so that the n-widths do not decay as rapidly as in the convection-diffusion case. Nevertheless, the rate-optimality still shows a relatively fast convergence for the reduced spaces X_n shown below.

The first example concerns (2.17) (with $\mu = s$ ranging over a quarter circle, $D = (0, 1)^2$) for $f_\circ \equiv 1$, $g \equiv 0$. In the second example, we take $f_\circ(x_1, x_2) = 0.5$ for $x_1 < x_2$, $f_\circ(x_1, x_2) = 1$ for $x_1 \geq x_2$ (Tables 2.2, 2.3 and Fig. 2.4).

Table 2.2 Numerical results for Example 1, maximal truth error in L_2 0.000109832

Dimension				Maximal error between		
Trial	Test	δ	Maximal surr	rb truth	rb L_2	Surr/err
4	11	3.95e−01	8.44e−03	2.45e−02	2.45e−02	3.45e−01
10	33	4.32e−01	3.37e−03	5.74e−03	5.74e−03	5.87e−01
16	57	4.32e−01	1.50e−03	2.56e−03	2.56e−03	5.84e−01
20	74	4.16e−01	1.21e−03	2.10e−03	2.10e−03	5.77e−01
24	91	4.05e−01	7.27e−04	1.58e−03	1.58e−03	4.61e−01

Table 2.3 Numerical results for Example 2 after a single cycle of iterative tightening. Maximal truth error in L_2 0.0154814

Dimension				Maximal error between		
Trial	Test	δ	Maximal surr	rb truth	rb L_2	Surr/err
First reduced basis creation						
20	81	3.73e−01	2.71e−02	5.46e−02	5.62e−02	4.82e−01
Second reduced basis creation						
10	87	3.51e−01	6.45e−02	7.40e−02	7.53e−02	8.57e−01
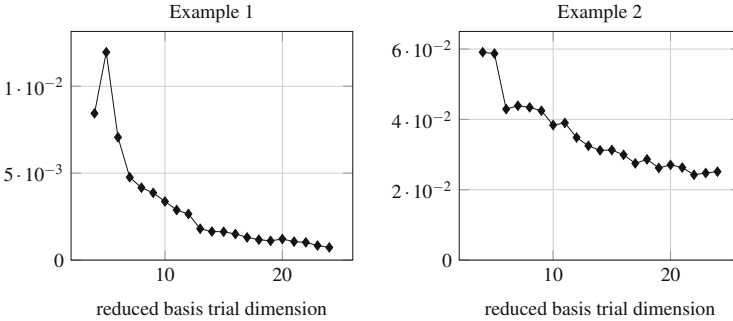
Fig. 2.4 Surrogates of the reduced basis approximation for Examples 1 and 2

2.5 Sparse Tensor Approximation for Radiative Transfer

We now extend the parametric transport problem (2.17) to the *radiative transport problem* (RTP) (see, e.g., [37]) which consists in finding the *radiative intensity* $u : D \times \mathcal{S} \to \mathbb{R}$, defined on the Cartesian product of a bounded physical domain $D \subset \mathbb{R}^d$, where $d = 2, 3$, and the unit $d_\mathbb{S}$-sphere as the parameter domain: $\mathcal{P} = \mathcal{S}$ with $d_\mathbb{S} = 1, 2$. Given an *absorption coefficient* $\kappa \geq 0$, a *scattering coefficient* $\sigma \geq 0$, and a *scattering kernel* or *scattering phase function* $\Phi > 0$, which is normalized to $\int_\mathcal{S} \Phi(s, s') \mathrm{d}s' = 1$ for each direction s, one defines the transport operator $\mathrm{T}u := (s \cdot \nabla_x + \kappa)u$, and the scattering operator $\mathrm{Q}u := \sigma \mathrm{Q}_1 u = \sigma(u - \int_\mathcal{S} \Phi(s, s')u(x, s')\mathrm{d}s')$. The radiative intensity is then given by

$$(\mathrm{T} + \mathrm{Q})u = f, \quad u|_{\partial\Omega_-} = g, \tag{2.48}$$

where $f := \kappa I_b$, $\partial\Omega_- := \{(x, s) \in \partial D \times \mathcal{S} : s \cdot n(x) < 0\}$, and g denote the source term, the inflow-boundary, and the inflow-boundary values, respectively. As before, $\Gamma_-(s) := \{x \in \partial D : s \cdot n(x) < 0\}$ (see (2.18)) stands for the physical inflow-boundary.

The partial differential equation (2.48) is known as *stationary monochromatic radiative transfer equation* (RTE) with scattering, and can be viewed as (nonlocal) extension of the parametric transport problem (2.17), where the major difference

to (2.17) is the scattering operator Q. Sources with support contained in D are modeled by the *blackbody intensity* $I_b \geq 0$, radiation from sources outside of the domain or from its enclosings is prescribed by the *boundary data* $g \geq 0$.

Deterministic numerical methods for the RTP which are commonly used in engineering comprise the *discrete ordinates* (S_N-) *method* and the *spherical harmonics* (P_N-) *method*.

In the *discrete ordinate method* (DOM), the angular domain is collocated by a finite number of fixed propagation directions in the angular parameter space; in this respect, the DOM resembles the greedy collocation in the parameter domain: each of the directions Eq. (2.48) results in a spatial PDE which is solved (possibly in parallel) by standard finite differences, finite elements, or finite volume methods.

In the *spherical harmonics method* (SHM), a spectral expansion with spatially variable coefficients is inserted as ansatz into the variational principle Eq. (2.48). By orthogonality relations, a coupled system of PDEs (whose type can change from hyperbolic to elliptic in the so-called diffuse radiation approximation) for the spatial coefficients is obtained, which is again solved by finite differences or finite elements.

The common deterministic methods S_N- and P_N-approximation exhibit the so-called "curse of dimensionality": the error with respect to the total numbers of degrees of freedom (DoF) M_D and $M_\mathscr{S}$ on the physical domain D and the parameter domain $\mathscr{S}$ scales with the dimension d and $d_\mathbb{S}$ as $O(M_D^{-s/d} + M_\mathscr{S}^{-t/d_\mathbb{S}})$ with positive constants s and t.

The so called *sparse grid approximation method* alleviates this curse of dimensionality for elliptic PDEs on cartesian product domains, see [7] and the references therein. Widmer et al. [41] has developed a sparse tensor method to overcome the curse of dimensionality for radiative transfer with a wavelet (isotropic) discretization of the angular domain. Under certain regularity assumptions on the absorption coefficient κ and the blackbody intensity I_b, their method achieves the typical benefits of sparse tensorization: a log-linear complexity in the number of degrees of freedom of a component domain with an essentially (up to a logarithmic factor) undeteriorated rate of convergence. However, scattering had not been addressed in that work.

In order to include scattering and to show that the concepts of sparse tensorization can also be applied to common solution methods, sparse tensor versions of the spherical harmonics approximation were developed extending the "direct sparse" approach by [41]. The presently developed version also accounts for scattering [27]. For this sparse spherical harmonics method, we proved that the benefits of sparse tensorization can indeed be harnessed.

As a second method a *sparse tensor product version of the DOM* based on the *sparse grid combination technique* was realized and analyzed in [26, 28]. Solutions to discretizations of varying discretization levels, for a number of collocated transport problems, and *with scattering discretized by combined Galerkin plus quadrature approximation in the transport collocation directions* are combined in this method to form a sparse tensor solution that we proved in [26, 28] breaks the curse of dimensionality as described above. These benefits hold as long as the exact solution of the RTE is sufficiently regular. An overview follows.

2.5.1 Sparse Discrete Ordinates Method (Sparse DOM)

We adopt a formulation where the inflow boundary conditions are enforced in a weak sense. To this end, we define the boundary form (see, e.g., [26])

$$\partial b(u, v) := (v, \mathbf{s} \cdot \mathbf{n}u)_{L^2(\partial\Omega_-)} = \int_{\mathscr{S}} \int_{\Gamma_-(\mathbf{s})} \mathbf{s} \cdot \mathbf{n}uv \mathrm{d}\mathbf{x} \mathrm{d}\mathbf{s} \ . \tag{2.49}$$

Writing for $v : D \times \mathscr{S} \to \mathbb{R}$ briefly $\|v\| := \|v\|_{L_2(D \times \mathscr{S})}$, the norms

$$\|v\|_-^2 := -\partial b(v, v), \quad \|v\|_1^2 := \|v\|^2 + \|\mathbf{s} \cdot \nabla_x v\|^2 + \|Q_1 v\|^2 + \|v\|_-^2$$

define the Hilbert space $\mathscr{V}_1 := \{v \in L_2(D \times \mathscr{S}) : \|v\|_1 < \infty\}$. The SUPG-stabilized Galerkin variational formulation reads: find $u \in \mathscr{V}_1$ such that

$$(Rv, (T + Q)u)_{L^2(D \times \mathscr{S})} - 2\partial b(u, v) = (Rv, f)_{L_2(D \times \mathscr{S})} - 2\partial b(g, v) \quad \forall v \in \mathscr{V}_1 \tag{2.50}$$

with SUPG stabilization $Rv := v + \eta \mathbf{s} \cdot \nabla_x v$, where $\eta \approx 2^{-L}$.

For the discretization of (2.50), we replace $\mathscr{V}_1$ by $V^{L,N} = V_D^L \otimes V_{\mathscr{S}}^N$. In the physical domain, standard P_1-FEM with a one-scale basis on a uniform mesh of width $h_{\max} \lesssim 2^{-L}$ is used, in the angular domain, piecewise constants on a quasiuniform mesh of width $h_{\max} \lesssim N^{-1}$. Fully discrete problems are obtained with a one-point quadrature in the angular domain. The resulting Galerkin formulation (2.50) can be shown to result in the same linear system of equations as the standard collocation discretization [26, Sec. 5.2]. The solution is constructed with the sparse grid combination technique (see [7]):

$$\hat{u}_{L,N} = \sum_{\ell_D=0}^{L} \left(u_{\ell_D, \ell_{\mathscr{S}}^{\max}(\ell_D)} - u_{\ell_D, \ell_{\mathscr{S}}^{\max}(\ell_D+1)} \right),$$

where $u_{\ell_D, \ell_{\mathscr{S}}} \in V^{\ell_D, \ell_{\mathscr{S}}}$ denotes the solution to a full tensor subproblem of physical resolution level ℓ_D and angular resolution level $\ell_{\mathscr{S}}$. The maximum angular index $\ell_{\mathscr{S}}^{\max} = 2^{\lfloor \log_2(N+1) \rfloor / L(L-l_D)}$ ensures that the angular resolution decreases when the physical resolution increases and vice versa.

While the full tensor solution $u_{L,N}$ requires $O(2^{dL}N^{d_{\mathbb{S}}})$ degrees of freedom, the sparse solution involves asymptotically at most $O((L+\log N)(2^{dL} + N^{d_{\mathbb{S}}}))$ degrees of freedom [26, Lemma 5.6]. At the same time, we have

$$\|u - u_{L,N}\|_1 \leq C 2^{-L} \|u\|_{H^{2,0}(D \times \mathscr{S})} + N^{-1} \|u\|_{H^{1,1}(D \times \mathscr{S})},$$

while for solutions in $H^{2,1}(D \times \mathscr{S}) \subset (H^{2,0}(D \times \mathscr{S}) \cap H^{1,1}(D \times \mathscr{S}))$

$$\|u - \hat{u}_{L,N}\|_1 \le CL \max\{2^{-L}, N^{-1}\}\|u\|_{H^{2,1}(D \times \mathscr{S})},$$

where C is an absolute constant.

2.5.2   Numerical Experiment

To evaluate the solution numerically we monitor the incident radiation $G(\boldsymbol{x}) = \int_{\mathscr{S}} u(\boldsymbol{x}, \boldsymbol{s}) \mathrm{d}\boldsymbol{s}$ and its relative error $err(G_{L,N})_X = \|G - G_{L,N}\|_X / \|G\|_X$, $X = L_2(D), H^1(D)$.

The setting for the experiment is $D = (0,1)^d$, $\mathscr{S} = \mathscr{S}_{\mathbb{S}}^d$. We solve the RTP (2.48) with isotropic scattering $\Phi(\boldsymbol{s}, \boldsymbol{s}') = 1/|\mathscr{S}|$ and with zero inflow boundary conditions $g = 0$. A blackbody radiation $I_b(\boldsymbol{x}, \boldsymbol{s})$ corresponding to the exact solution

$$u(\boldsymbol{x}, \boldsymbol{s}) = \frac{3}{16\pi}(1 + (\boldsymbol{s} \cdot \boldsymbol{s}')^2) \prod_{i=1}^{3}(-4x_i(x_i - 1)),$$

with fixed $\boldsymbol{s}' = (1/\sqrt{3}, 1/\sqrt{3}, 1/\sqrt{3})^\top$ is inserted in the right hand side functional in (2.50). The absorption coefficient is set to $\kappa = 1$, the scattering coefficient to $\sigma = 0.5$.

This $3 + 2$-dimensional problem was solved with a parallel C++ solver designed for the sparse tensor solution of large-scale radiative transfer problems. Figure 2.5 shows the superior efficiency of the sparse approach with respect to number of degrees of freedom vs. achieved error. The convergence rates indicate that the curse
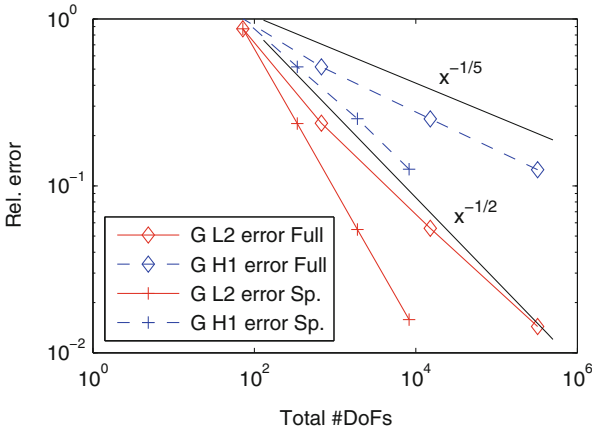
Fig. 2.5 Convergence in incident radiation with full and sparse DOM. Resolution for reference solution was $L_{\mathrm{ref}} = 4$. Reference slopes provided as visual aids only

of dimensionality is mitigated by the sparse DOM. Further gains are expected once the present, nonadaptive sparse DOM is replaced by the greedy versions outlined in Sect. 2.3.1.1.

References

1. Alpert, B.K.: A class of bases in L^2 for the sparse representation of integral operators. SIAM J. Math. Anal. **24**, 246–262 (1993)
2. Bachmayer, M., Dahmen, W.: Adaptive near-optimal rank tensor approximation for high-dimensional operator equations. Found. Comput. Math. March, 2014, doi: 10.1007/s10208-013-9187-3, http://arxiv.org/submit/851475
3. Barrett, J.W., Morton, K.W.: Approximate symmetrization and Petrov-Galerkin methods for diffusion-convection problems. Comput. Methods Appl. M. **45**, 97–12 (1984)
4. Binev, P., Cohen, A., Dahmen, W., DeVore, R., Petrova, G., Wojtaszczyk, P.: Convergence rates for greedy algorithms in reduced basis methods. SIAM J. Math. Anal. **43**, 1457–1472 (2011)
5. Bochev, P.B., Gunzburger, M.D.: Least-Squares Finite Element Methods. Applied Mathematical Sciences. Springer, New York (2009)
6. Buffa, A., Maday, Y., Patera, A.T., Prud'homme, C., Turinici, G.: A priori convergence of the greedy algorithm for the parameterized reduced basis. ESAIM: Math. Model. Numer. Anal. **46**, 595–603 (2012)
7. Bungartz, H.-J., Griebel, M.: Sparse grids. In: Iserles, A. (ed.) Acta Numerica, vol. 13, pp. 147–269. Cambridge University Press, Cambridge (UK) (2004)
8. Cai, Z., Manteuffel, T., McCormick, S., Ruge, J.: First-order system $\mathscr{L}\mathscr{L}^*$ ($FOSLL$)*: scalar elliptic partial differential equtions. SIAM J. Numer. Anal. **39**, 1418–1445 (2001)
9. Candès, E.J., Donoho, D.L.: New tight frames of curvelets and optimal representations of objects with piecewise-C^2 singularities. Commun. Pure Appl. Math. **57**, 219–266 (2002)
10. Chen, L., Sun, P., Xu, J.: Optimal anisotropic simplicial meshes for minimizing interpolation errors in L^p norm. Math. Comput. **76**, 179–204 (2007)
11. Chkifa, A., Cohen, A., DeVore, R., Schwab, C.: Sparse adaptive Taylor approximation algorithms for parametric and stochastic elliptic PDEs. ESAIM: Math. Model. Numer. Anal. **47**(1), 253–280 (2013). http://dx.doi.org/10.1051/m2an/2012027
12. Cohen, A., Mirebeau, J.-M.: Greedy bisection generates optimally adapted triangulations. Math. Comput. **81**, 811–837 (2012)
13. Cohen, A., DeVore, R., Schwab, C.: Convergence rates of best N-term Galerkin approximations for a class of elliptic sPDEs. Found. Comput. Math. **10**, 615–646 (2010)
14. Cohen, A., Dahmen, W., Welper, G.: Adaptivity and variational stabilization for convection-diffusion equations. ESAIM: Math. Model. Numer. Anal. **46**, 1247–1273 (2012)
15. Cohen, A., Dyn, N., Hecht, F., Mirebeau, J.-M.: Adaptive multiresolution analysis based on anisotropic triangulations. Math. Comput. **81**, 789–810 (2012)
16. Dahmen, W., Huang, C., Schwab, C., Welper, G.: Adaptive Petrov-Galerkin methods for first order transport equations. SIAM J. Numer. Anal. **50**, 2420–2445 (2012)
17. Dahmen, W., Kutyniok, G., Lim, W.-Q, Schwab, C., Welper, G.: Adaptive anisotropic discretizations for transport equations, preprint, 2014
18. Dahmen, W., Plesken, C., Welper, G.: Double greedy algorithms: reduced basis methods for transport dominated problems. ESAIM: Math. Model. Numer. Anal. **48**, 623–663 (2014). doi:10.1051/m2an/2013103, http://arxiv.org/abs/1302.5072
19. Demkowicz, L.F., Gopalakrishnan, J.: A class of discontinuous Petrov-Galerkin methods I: the transport equation. Comput. Methods Appl. Mech. Eng. **199**, 1558–1572 (2010)
20. Demkowicz, L., Gopalakrishnan, J.: A class of discontinuous Petrov-Galerkin methods. Part II: optimal test functions. Numer. Methods Part. Differ. Equ. **27**, 70–105 (2011)

21. DeVore, R., Petrova, G., Wojtaszczyk, P.: Greedy algorithms for reduced bases in Banach spaces. Constr. Approx. **37**, 455–466 (2013)
22. Dolejsi, V.: Anisotropic mesh adaptation for finite volume and finite element methods on triangular meshes. Comput. Vis. Sci. **1**, 165–178 (1998)
23. Donoho, D.L.: Sparse components of images and optimal atomic decompositions. Constr. Approx. **17**, 353–382 (2001)
24. Donoho, D.L., Kutyniok, G.: Microlocal analysis of the geometric separation problem. Commun. Pure Appl. Math. **66**, 1–47 (2013)
25. Gerner, A., Veroy-Grepl, K.: Certified reduced basis methods for parametrized saddle point problems. SIAM J. Sci. Comput. **35**, 2812–2836 (2012)
26. Grella, K.: Sparse tensor approximation for radiative transport. PhD thesis 21388, ETH Zurich (2013)
27. Grella, K., Schwab, C.: Sparse tensor spherical harmonics approximation in radiative transfer. J. Comput. Phys. **230**, 8452–8473 (2011)
28. Grella, K., Schwab, C.: Sparse discrete ordinates method in radiative transfer. Comput. Methods Appl. Math. **11**, 305–326 (2011)
29. King, E.J., Kutyniok, G., Zhuang, X.: Analysis of inpainting via clustered sparsity and microlocal analysis. J. Math. Imaging Vis. **48**, 205–234 (2014)
30. Kittipoom, P., Kutyniok, G., Lim, W.-Q: Construction of compactly supported shearlet frames. Constr. Approx. **35**, 21–72 (2012)
31. Kutyniok, G., Labate, D.: Resolution of the wavefront set using continuous shearlets. Trans. Am. Math. Soc. **361**, 2719–2754 (2009)
32. Kutyniok, G., Labate, D.: Shearlets: Multiscale Analysis for Multivariate Data. Birkhäuser, Boston (2012)
33. Kutyniok, G., Lim, W.-Q: Compactly supported shearlets are optimally sparse. J. Approx. Theory **163**, 1564–1589 (2011)
34. Kutyniok, G., Lemvig, J., Lim, W.-Q: Optimally sparse approximations of 3D functions by compactly supported shearlet frames. SIAM J. Math. Anal. **44**, 2962–3017 (2012)
35. Lim, W.-Q: The discrete shearlet transform: a new directional transform and compactly supported shearlet frames. IEEE Trans. Image Proc. **19**, 1166–1180 (2010)
36. Mirebeau, J.-M.: Adaptive and anisotropic finite element approximation: theory and algorithms, PhD thesis, Université Pierre et Marie Curie – Paris VI (2011). http://tel.archives-ouvertes.fr/tel-00544243
37. Modest, M.F.: Radiative Heat Transfer, 2nd edn. Elsevier, Amsterdam (2003)
38. Rozza, G., Veroy, K.: On the stability of reduced basis techniques for Stokes equations in parametrized domains. Comput. Methods Appl. Mech. **196**, 1244–1260 (2007)
39. Sen, S., Veroy, K., Huynh, D.B.P., Deparis, S., Nguyn, N.C., Patera, A.T.: "Natural norm" a-posteriori error estimators for reduced basis approximations. J. Comput. Phys. **217**, 37–62 (2006)
40. Welper, G.: Infinite dimensional stabilization of convection-dominated problems. PhD thesis, RWTH Aachen (2013). http://darwin.bth.rwth-aachen.de/opus3/volltexte/2013/4535/
41. Widmer, G., Hiptmair, R., Schwab, C.: Sparse adaptive finite elements for radiative transfer. J. Comput. Phys. **227**, 6071–6105 (2008)