

Preconditioners for Krylov subspace methods: An overview

John W. Pearson¹ | Jennifer Pestana²

¹School of Mathematics, The University of Edinburgh, Edinburgh, UK

²Department of Mathematics and Statistics, University of Strathclyde, Glasgow, UK

Correspondence

Jennifer Pestana, Department of Mathematics and Statistics, University of Strathclyde, 16 Richmond Street, Glasgow G1 1XQ, UK. Email: jennifer.pestana@strath.ac.uk

Abstract

When simulating a mechanism from science or engineering, or an industrial process, one is frequently required to construct a mathematical model, and then resolve this model numerically. If accurate numerical solutions are necessary or desirable, this can involve solving large-scale systems of equations. One major class of solution methods is that of preconditioned iterative methods, involving preconditioners which are computationally cheap to apply while also capturing information contained in the linear system. In this article, we give a short survey of the field of preconditioning. We introduce a range of preconditioners for partial differential equations, followed by optimization problems, before discussing preconditioners constructed with less standard objectives in mind.

KEYWORDS

iterative method, Krylov subspace method, optimization, partial differential equations, preconditioning

1 | INTRODUCTION

Solving large linear systems $Ax = b$, where $A \in \mathbb{C}^{n \times n}$ and $b \in \mathbb{C}^n$, is a mainstay of scientific computing. Such systems arise in many applications—perhaps most frequently the discretization of partial differential equations (PDEs) and optimization problems—and their solution can be extremely time-consuming. Moreover, for certain time-dependent, nonlinear, or inverse problems, for instance, multiple linear systems must be solved, so a reduction in the linear solve time results in a huge increase in the efficiency of the overall numerical scheme.

For the solution of a wide class of linear systems, the use of *direct methods* is both popular and powerful (see [1] for a survey of such methods for sparse systems). Here, a sequence of operations is performed on the matrix in order to generate a suitable factorization, therefore returning a single estimate of the solution of the system. However, if the matrix in question becomes sufficiently large for a particular computer architecture, the storage and operation costs of a direct method may become excessive. In such a case it becomes valuable to apply *iterative methods*, where no such factorization is required, in order to generate a sequence of approximate solutions. Such methods generally require matrix-vector multiplications at each iteration, but these may be performed without storing the entire matrix: one may work with individual blocks of the matrix, or indeed apply “matrix-free” approaches (see [2,3] for instance). Here, we focus specifically on Krylov subspace methods, which we detail in Section 2; for such methods A is indeed only accessed via matrix-vector multiplications. However, without the use of preconditioning (that we describe in greater detail in Section 2) the Krylov solver may not converge, or may take many iterations to return an acceptably accurate solution. In the best case, *preconditioned iterative methods* can substantially reduce the storage and operation costs as opposed to direct methods, while achieving greatly improved convergence properties that may often be proved a priori.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2020 The Authors. *GAMM - Mitteilungen* published by Wiley-VCH GmbH on behalf of Gesellschaft für Angewandte Mathematik und Mechanik.

Two fundamental goals when devising a preconditioner can therefore be summarized as follows:

- Facilitating rapid convergence of an iterative method.
- Guaranteeing a reduction in terms of storage and computational operations as opposed to direct methods.

Accordingly, preconditioning allows for the solution of problems that cannot be tackled using either direct schemes or unpreconditioned iterative methods.

The design of such a preconditioner necessitates a trade-off, as one then needs to solve a linear system involving the preconditioner at least once per iteration. Hence, a preconditioner must be constructed such that it is cheap to apply its inverse operation to a given vector, while also capturing the characteristics of the original system in some sense. The second criterion will be explored further throughout this paper, as its interpretation depends on several factors including which iterative method is used.

There are already several excellent surveys of preconditioners in general, and for specific applications; we hope that this article complements these existing works. In particular, we mention the extensive surveys by Benzi [4] and Wathen [5], as well as the books by Bertaccini and Durastante [6] and Chen [7]. Further works surveying certain areas of preconditioning include [8-11]. A key structural difference in this paper is that we predominantly categorize by application area rather than by type of preconditioner. Throughout this text, we additionally point to surveys on individual topics.

We begin our discussion by describing the most widely used Krylov subspace methods in Section 2. We particularly emphasize the convergence theory available for these methods, since this can be used to determine preconditioners that fulfill the aims outlined above. Section 3 then details preconditioners appropriate for PDE problems, while Section 4 discusses preconditioners for linear systems that arise in optimization. Although preconditioning typically focuses on ensuring that the preconditioned matrix has “favorable” properties, there are also other ways to precondition, and reasons for doing so, as we illustrate through selected examples in Section 5. Finally, we summarize our key messages in Section 6.

Notation: Throughout, we use bold lowercase letters, for example \mathbf{x} , to represent vectors, and uppercase letters, for example M , for matrices. We denote by A a matrix that is being solved for and by P a preconditioner, except in the case that the block structure is important, when we use the calligraphic letters \mathcal{A} and \mathcal{P} . In this case A can denote a sub-block of \mathcal{A} . A general matrix is represented by M . We denote the Hermitian transpose of \mathbf{x} by \mathbf{x}^H and the transpose by \mathbf{x}^T . If $M \in \mathbb{C}^{n \times n}$ is Hermitian positive definite then it defines a norm, which we denote by $\|\mathbf{x}\|_M := \sqrt{\mathbf{x}^H M \mathbf{x}}$. We will call a preconditioner *optimal* if it requires $O(n)$ floating point operations to apply, and if the number of iterations required to attain a certain accuracy can be bounded independently of n . Whenever we cite several references consecutively, such as [4,5,7], this is a possibly nonexhaustive list and should be read “see, for example, [4,5,7] and the references therein”.

2 | KRYLOV SUBSPACE METHODS

Preconditioning has a surprisingly long history, with a crucial early reference to this term made by Alan Turing [12]. Another milestone was the paper on incomplete Cholesky preconditioners by Meijerink and van der Vorst [13], which clearly illustrated the potential of preconditioning to improve the efficiency of iterative solvers. In this section, to better understand the role of preconditioners, we provide a short introduction to Krylov subspace methods. This is by no means comprehensive, and we refer the interested reader to the excellent books by Greenbaum [14], Liesen and Strakoš [15], and Saad [16], and surveys by Eiermann and Ernst [17], Liesen and Tichý [18], and Simoncini and Szyld [19]. We note that although our focus here is on Krylov methods, the approaches to preconditioning described in this article can certainly be applied to other iterative algorithms.

Stationary iterative methods have been used at least since the time of Gauss, but over the last few decades they have been largely superseded by Krylov subspace methods, which approximate \mathbf{x} from Krylov subspaces. Named after Aleksei Nikolaevich Krylov, who used these spaces to analyze oscillations of mechanical systems [20], Krylov subspaces take the form

$$\mathcal{K}_k(A, \mathbf{v}) := \text{span}\{\mathbf{v}, A\mathbf{v}, \dots, A^{k-1}\mathbf{v}\},$$

where $A \in \mathbb{C}^{n \times n}$ and $\mathbf{v} \in \mathbb{C}^n$.

The first Krylov subspace method for linear systems, namely the conjugate gradient (CG) method, was proposed by Hestenes and Stiefel [21], with Lanczos also working on closely related topics in this period [22,23]. The CG method must terminate in at most n iterations in exact arithmetic. Thus, although Hestenes and Stiefel [21] and Lanczos [23] refer to CG and the related Lanczos method as iterative solvers, over the next two decades mathematicians (in contrast to engineers and physicists) considered CG a direct method. However, in this capacity it was often outperformed by other approaches. It was only after John Reid's observation [24] that good approximations to \mathbf{x} could be obtained in far fewer than n steps, that Krylov methods were again viewed as iterative methods.¹ Renewed interest followed, with new methods developed and analyzed. For Hermitian matrices the most popular are probably MINRES and SYMMLQ [26] while for non-Hermitian matrices such methods include GMRES [27], FOM [28], LSQR [29], LSMR [30], and families of methods based on QMR [31], BiCG [32] and BiCGStab [33], and IDR [34,35].

Krylov subspace methods for linear systems start from an initial guess \mathbf{x}_0 (often the zero vector), compute the initial residual $\mathbf{r}_0 = \mathbf{b} - A\mathbf{x}_0$, and then (typically) select iterates $\mathbf{x}_1, \mathbf{x}_2, \dots$ such that

$$\mathbf{x}_k - \mathbf{x}_0 \in \mathcal{K}_k(A, \mathbf{r}_0), \quad (1)$$

thereby projecting the original problem to one of much smaller dimension. This approach is naturally linked to polynomial approximation, since $\mathbf{x}_k - \mathbf{x}_0 \in \mathcal{K}_k(A, \mathbf{r}_0) \Leftrightarrow \mathbf{x}_k = \mathbf{x}_0 + q_{k-1}(A)\mathbf{r}_0$ for some $q_{k-1} \in \Pi_{k-1}$ with Π_k denoting the set of polynomials of degree at most k . Accordingly, the k th residual satisfies

$$\mathbf{r}_k = \mathbf{b} - A\mathbf{x}_k = (I - Aq_{k-1}(A))\mathbf{r}_0 = p_k(A)\mathbf{r}_0, \quad p_k \in \Pi_k, \quad p(0) = 1.$$

We have not yet described how iterates are chosen from these Krylov subspaces, and indeed this choice distinguishes different methods. We will not discuss this in detail here, but will instead focus on the available convergence theory for these methods.

2.1 | Hermitian problems

We first consider the CG method for Hermitian positive definite A , which is perhaps the best understood Krylov method. CG chooses \mathbf{x}_k according to (1) such that $\|\mathbf{e}_k\|_A$ is minimized, where $\mathbf{e}_k = \mathbf{x} - \mathbf{x}_k$. It requires only short-term recurrences so that the computational work at each iteration is fixed. Crucially for preconditioning we can bound $\|\mathbf{e}_k\|_A$ as follows:

$$\frac{\|\mathbf{e}_k\|_A}{\|\mathbf{e}_0\|_A} \leq \min_{\substack{p \in \Pi_k \\ p(0)=1}} \max_{\lambda \in \sigma(A)} |p(\lambda)|,$$

where $\sigma(A)$ is the spectrum of A . This bound is sharp, in the sense that for every k there is a (possibly different) initial guess \mathbf{x}_0 for which it is attained. That said, given a specific initial residual, that is, a specific right-hand side and initial guess, the convergence rate may be much faster than is predicted by the above bound. We refer to the book by Liesen and Strakoš [15, section 5.6] and the references therein for a careful CG convergence analysis.

When A is Hermitian but indefinite, we can instead apply MINRES, which also has fixed work per step and for which an analogous convergence bound can be derived. MINRES chooses \mathbf{x}_k such that $\|\mathbf{r}_k\|_2$ is minimized subject to (1), and in this case

$$\frac{\|\mathbf{r}_k\|_2}{\|\mathbf{r}_0\|_2} \leq \min_{\substack{p \in \Pi_k \\ p(0)=1}} \max_{\lambda \in \sigma(A)} |p(\lambda)|. \quad (2)$$

This bound is descriptive in the same sense as the CG bound mentioned above [18], but the eigenvalues of A may also be negative.

We see from the above bounds that convergence of CG and MINRES is strongly related to the spectral properties of A . Since these are governed by the problem at hand, it is often desirable to improve the convergence rate by solving an

¹This was also noted in [25], but its potential was not recognized until later.

equivalent linear system with “better” spectral properties, that is, by preconditioning. To preserve symmetry, for CG and MINRES we precondition symmetrically, that is, one may instead consider the formalism

$$C^{-1}AC^{-H}\mathbf{y} = C^{-1}\mathbf{b}, \quad \mathbf{x} = C^{-H}\mathbf{y},$$

where $C \in \mathbb{C}^{n \times n}$ is nonsingular. For this preconditioned system, the spectral properties of $C^{-1}AC^{-H}$ become important, and by a careful choice of C it is hoped that convergence will be faster. We note that by rewriting the CG or MINRES algorithm it is possible to avoid computations with C and C^H and work only with the symmetric positive definite matrix $P = CC^H$. However, the preconditioner for CG and MINRES must be Hermitian positive definite in general.

2.2 | Non-Hermitian problems

GMRES is an extension of MINRES to non-Hermitian systems, and because it minimizes $\|\mathbf{r}_k\|_2$ it has the most complete convergence theory out of solvers for such systems. If A is normal (2) holds, although now the eigenvalues of A may be complex, and this makes the convergence of GMRES more challenging to understand. However, when A is far from normal, no convergence bound is universally descriptive [36], and the spectrum alone cannot characterize the convergence rate of GMRES [37-40]. A major disadvantage of GMRES is that the computational work grows with each iteration, and one option is to restart GMRES [16, chapter 6]. Alternatively we may use a method based on bi-orthogonalization, such as QMR, BiCGStab, or IDR. However, none of these methods has a tractable optimality property, and so convergence results are extremely limited.

Thus for non-Hermitian, particularly nonnormal, problems preconditioning is heuristically motivated although, as we shall describe, this has not prevented the development of many effective preconditioners for a range of such systems. We note that for non-Hermitian problems it is possible to perform

- left preconditioning, that is, solve $P^{-1}A\mathbf{x} = P^{-1}\mathbf{b}$;
- right preconditioning, that is, solve $AP^{-1}\mathbf{y} = \mathbf{b}$, with $\mathbf{x} = P^{-1}\mathbf{y}$;
- or split preconditioning, that is, solve $P_1^{-1}AP_2^{-1}\mathbf{y} = P_1^{-1}\mathbf{b}$, with $\mathbf{x} = P_2^{-1}\mathbf{y}$.

A key restriction is that the preconditioner must represent some linear operator. If this is not the case a flexible Krylov subspace should be used [16, section 9.4].

We end this section by remarking that it is always possible to symmetrize a problem by, for example, solving the normal equations $A^H A \mathbf{x} = A^H \mathbf{b}$. CG or MINRES can be applied to this symmetrized system, although in practice the more numerically stable but mathematically equivalent LSQR (for CG) or LSMR (for MINRES) are used. In this case we again recover descriptive convergence theory. On the other hand, convergence rates for problems solved using the normal equations are often significantly worse than for the original linear system, and preconditioning $A^H A$ (or AA^H) is often challenging; indeed it is possible that $P^H P$ is an arbitrarily poor preconditioner for $A^H A$ even if P is a suitable preconditioner for A [41].

3 | PRECONDITIONERS FOR PDES AND RELATED PROBLEMS

Often, linear systems arise upon the discretization of an infinite-dimensional problem, for example differential equations and PDEs, or integral equations. Here we focus on second-order PDEs, but some of the underlying themes and principles can be applied to, or adapted for, discretizations of other infinite-dimensional problems.

An important point is that solving the linear system is not an end in itself; instead, the overarching goal is to approximate the solution to the infinite-dimensional problem. The accuracy of this solution is affected by the discretization error as well as the error in terminating the iterative solver (the algebraic error). The interplay between these errors [42,43], and their effect on the stopping criteria [44-46], should be considered, although this is typically a nontrivial task.

The convergence rates of iterative methods for discretized problems are influenced both by the properties of the infinite-dimensional problem and the choice of discretization. Knowledge of the properties of the underlying PDE (particularly the weak form) has been exploited to develop so-called “operator” preconditioners [47-52] [53, chapter 2]; for an overview of these, see [54, chapter 4] [55]. The recent paper [56] also treats operator preconditioners in infinite- and

finite-dimensional settings and describes an abstract framework for operator preconditioners in terms of a splitting of a Hilbert space into a finite number of subspaces. Other preconditioners that, at least in their original forms, are based on knowledge of the underlying differential equation include multigrid and domain decomposition, which we discuss in more detail below.

Discretization choices also affect the properties of the coefficient matrix, and hence the convergence rate of the Krylov subspace method. In particular, PDE problems require global transfer of information but popular discretizations, for example finite element, finite difference, and finite volume methods, lead to sparse matrices that transfer information only locally. Preconditioning PDE problems can be viewed as transforming to a more appropriate basis [54, chapter 8]. This perspective has also been used more explicitly in hierarchical basis preconditioning [57,58] (see also the survey by Xu [59]), and cardinal basis preconditioning for radial basis discretizations [60,61], which were originally introduced in the context of interpolation [62].

In this section we discuss preconditioners for some prototypical PDEs; these PDEs either arise on their own or, increasingly commonly, as key components of more complicated, possibly nonlinear PDE problems. These include the Poisson, Helmholtz, and convection-diffusion equations, which represent problems involving a single scalar-valued variable, and the Stokes and Navier-Stokes equations, which are examples of systems of PDEs. We also discuss specialized methods for Toeplitz matrices, time-dependent problems, and problems with heterogeneous coefficients.

3.1 | Single PDEs

3.1.1 | Poisson problems

The first equation we consider is the generalized Poisson problem: find u such that

$$-\nabla \cdot (\alpha \nabla u) = f \quad \text{in } \Omega \quad (3)$$

subject to appropriate boundary conditions, where α is a sufficiently smooth positive function that is both bounded above and uniformly bounded away from zero on Ω , and $\Omega \subset \mathbb{R}^d$, $d \in \{2, 3\}$, is a bounded connected domain with piecewise smooth boundary $\partial\Omega$. (The standard Poisson problem is recovered when $\alpha \equiv 1$.) Equation (3) is elliptic and self-adjoint, and arises in many applications in, for example, chemistry, astrophysics, fluid dynamics, solid mechanics, electromagnetics, and image processing. It is also a subproblem of more complicated PDEs, such as the Navier-Stokes equations. Discretization by standard numerical methods leads to a linear system $A\mathbf{x} = \mathbf{b}$, where $A \in \mathbb{R}^{n \times n}$ is symmetric and at least positive semidefinite.

When (3) has separable coefficients on rectangular domains, it may be discretized by finite difference methods and then solved by tailored direct methods known as fast Poisson solvers [63-66]. However, for more complicated problems, domains, meshes, or discretizations, the preconditioned CG method is usually the method of choice. In these cases, fast Poisson solvers may be used as preconditioners, although their effectiveness deteriorates as the complexity of the problem increases. These fast Poisson solver preconditioners can also be viewed as operator preconditioners that approximate $(-\nabla^2)^{-1}$, see, for example, [55]. Recently, it was shown that when Laplace preconditioners are applied to finite element discretizations of (3) with nonconstant α , the eigenvalues of the preconditioned matrix are approximated by nodal values of α [67].

Multigrid

More robust preconditioners for elliptic, and indeed hyperbolic and parabolic, PDEs include multigrid methods, which are iterative solvers in their own right. Standard multigrid methods are based on the observation that when stationary iterations such as Jacobi or Gauss-Seidel are applied to problems from PDEs, in many cases they quickly reduce high-frequency components of the error but are less effective at removing smooth, low-frequency, components. Multigrid exploits this, and a sequence of smaller linear systems, to accelerate the convergence of these stationary iterations.

Multigrid is more easily understood by first considering a two-grid scheme. Within this two-grid iteration, a few steps of a stationary iterative method result in the approximation $\hat{\mathbf{x}}$ to \mathbf{x} , with smooth error $\mathbf{e} = \mathbf{x} - \hat{\mathbf{x}}$. We then improve $\hat{\mathbf{x}}$ by approximately solving the associated residual equation $A\mathbf{e} = \mathbf{r}$. Specifically, since \mathbf{e} , hence \mathbf{r} , are smooth they are well represented by a smaller (coarse grid) problem $A_c \mathbf{e}_c = \mathbf{r}_c$ that is more easily solved. Projecting \mathbf{e}_c back to the fine grid then

gives $\hat{\mathbf{e}} \approx \mathbf{e}$, and we obtain the improved approximation $\mathbf{x} \approx \hat{\mathbf{x}} + \hat{\mathbf{e}}$. Applying the stationary iterative method again at this stage (to post-smooth) is optional.

In multigrid, the process of smoothing and projecting is repeated recursively until we reach a problem that is small enough to be solved by a direct method; to obtain $\hat{\mathbf{e}}$ we then interpolate this approximate solution on successively finer grids by traversing the grid hierarchy in the opposite direction. The motivation for this repeated grid transfer is that low-frequency error components for the original problem become high-frequency components when represented on this coarse grid, which can again be smoothed by the stationary iteration. If a single two-grid iteration is performed at each recursive level we obtain a V-cycle; if two such iterations are performed on each recursive level we instead obtain a W-cycle, see, for example, [68, section 2.4].

Multigrid ideas appeared in early papers [69,70], but the seminal work by Brandt [71] is widely regarded as the foundation of modern multigrid methods. These solvers were geometric multigrid (GMG) methods, using a sequence of meshes for Ω . GMG is convenient for simpler problems and structured grids for which interpolation and restriction of the error are more straightforward.

For complex domains, unstructured meshes, and/or heterogeneous coefficients, it can be challenging to develop effective GMG methods. To overcome this, algebraic multigrid (AMG) methods were developed from the 1980s onwards [72-76]. The key distinction is that coarsening is performed automatically using only matrix entries, in a way that mimics GMG methods. AMG is now widely used, not only for problems that are challenging for GMG, but also to avoid programming problem-specific solvers.

Multigrid methods have been the subject of much research over recent decades and further details can be found in, for example, [68,77-80]. (Linear) multigrid is itself a (complicated) stationary iterative method; for many simpler problems it is also an optimal solver and in these cases it can be used on its own. For more complicated problems multigrid may converge slowly, if at all, because some error modes are not adequately damped. In this situation it is much more effective to use a fixed number of multigrid iterations as a preconditioner for the CG method, provided the multigrid method represents a symmetric positive definite preconditioner, for example a V-cycle or W-cycle with smoothing during the interpolation phase that is symmetric with respect to the smoothing used when coarsening. Note that using a variable number of multigrid iterations, or an otherwise nonlinear method, results in a nonlinear preconditioner that should not be used with standard Krylov subspace methods.

Domain decomposition

Domain decomposition methods also replace the original Poisson linear system by smaller problems but, in contrast to multigrid methods, they divide the domain into subregions and solve (3) on each. Doing so requires the imposition of artificial boundary conditions, called transmission conditions, on subdomain boundaries (except where the subdomain boundary is part of $\partial\Omega$). These transmission conditions allow the transfer of information between subdomains. While subdomain solves originally corresponded to physical regions of Ω , it is also possible to specify subdomains using subsets of rows and columns of A .

The first domain decomposition method was proposed by Schwarz in 1870 [81] to prove existence and uniqueness of solutions to (3) on complicated domains by dividing them into two simpler subdomains. This alternating Schwarz method solved over each subdomain in turn, using the most up-to-date solution information to impose Dirichlet boundary conditions. Multiplicative Schwarz methods generalize this idea, but additive Schwarz and related methods [82,83] are generally preferred; one reason for this is that they can be more easily parallelized. In particular, the popular restricted additive Schwarz method adds the solution to each overlapping subdomain problem, using a partition of unity in overlap regions.

Additive Schwarz, multiplicative Schwarz, and related methods are also stationary iterative schemes that can be used as standalone solvers provided they converge [84,85]. However, local transfer of information means that their convergence rates can be slow, and so they are usually supplemented by some sort of global coarse grid solve. This global solve may additionally deal with modes that impede convergence [86-88].

Like multigrid, for complex problems domain decomposition methods are often much more effective as preconditioners. There are many such preconditioners available that differ in their choice of subdomains, overlap, transmission conditions, and coarse grid solves. Domain decomposition methods remain an active research area, and we recommend several books and surveys [89-95].

3.1.2 | Helmholtz problems

A seemingly small modification of the Poisson equation (3) gives rise to the Helmholtz problem

$$-\nabla^2 u - k^2 u = f \quad \text{in } \Omega, \quad (4)$$

again augmented with appropriate boundary conditions, with k real-valued and $\Omega \subset \mathbb{R}^d$, $d \in \{2, 3\}$, a bounded connected domain with piecewise smooth boundary $\partial\Omega$. This important PDE features in, for example, wave propagation problems and quantum mechanics, with oscillatory problems—that are characterized by moderate to large values of k —typically of interest.

Once discretized, (4) leads to a linear system $A\mathbf{x} = \mathbf{b}$, where $A = K - k^2 M \in \mathbb{R}^{n \times n}$, with K the discretized negative Laplacian and M , for example, the identity matrix or a finite element mass matrix. A number of issues make the solution of this system challenging. First, complicated boundary conditions are often required for well-posedness and/or to mimic an infinite domain, and these can make A non-Hermitian, and possibly complex. Second, fine meshes are usually required to resolve the oscillatory solution and to avoid the “pollution effect”, that is, numerical dispersion; this results in extremely large linear systems. Finally, even if A is Hermitian, as we consider here, it is almost always indefinite. In fact, A is singular if k^2 is an eigenvalue of $M^{-1}K$, that is, in the presence of resonances, although MINRES, for example, can be applied to singular but consistent systems.

Standard preconditioners for elliptic problems generally perform poorly on discretizations of the indefinite Helmholtz problem. Multigrid preconditioners may struggle to smooth errors, and the oscillatory nature of the problem limits the sizes of coarse grids [96]; however, tailored multigrid [97-99] and multilevel [100] methods attempt to remedy these issues. Domain decomposition methods also require careful treatment, for example a sufficiently fine coarse grid [101], special coarse grid solves [102,103], and/or modified transmission conditions [104-106].

Instead of approximating A directly, it is possible to approximate a “nicer” elliptic operator such as the negative Laplace operator [107] or the shifted operator [108] $-\nabla^2 + k^2 I$ by standard multilevel approximations. Unfortunately, the performance of these preconditioners deteriorates as k increases, as might be expected. More general “shifted Laplace” preconditioners based on $-\nabla^2 + \gamma I$, with γ complex, typically perform better [109-111] but an appropriate choice of shift is important for efficiency [112,113]; small shifts reduce the number of Krylov subspace iterations but make the preconditioner itself difficult to approximate well. An additional issue is that a complex shift makes the problem complex and non-Hermitian. It is possible to instead solve a real block system, but this doubles the problem dimension and two elliptic subproblems must be solved per iteration [114].

Sweeping preconditioners have also received much attention of late. These incomplete (block) factorization preconditioners sweep through the variables in layers, using information about the underlying PDE to approximate Schur complements [115,116] (see the survey [117]). Recently, a different approach was proposed that utilizes contour integrals to project the spectrum of A , transforming the original problem to one that is easier to solve [118]. Surveys of Helmholtz preconditioners can be found in [119,120].

The key point is that the preconditioners of choice for the Helmholtz problem are not necessarily the same as for the Poisson problem. This is essentially because features of the PDE (oscillations, and loss of ellipticity) strongly influence properties of A , which must then be accounted for when designing preconditioners.

3.1.3 | Convection-diffusion problems

Non-self-adjoint PDEs lead to linear systems with non-Hermitian coefficient matrices that must be solved by one of the nonsymmetric Krylov methods introduced in Section 2. Among the best-studied PDEs of this type is the convection-diffusion equation

$$-\nabla^2 u + \mathbf{w} \cdot \nabla u = f$$

in $\Omega \subset \mathbb{R}^d$, $d \in \{2, 3\}$, with appropriate boundary conditions. One reason for this interest is that the Krylov subspace residual norms $\|\mathbf{r}_k\|_2$ for unpreconditioned linear systems arising from convection-diffusion problems are often characterized by an initial period of stagnation, during which the residual norm reduction is extremely slow, followed by a second, faster convergence phase. This two-phase behavior is challenging to characterize using convergence bounds, and may be

poorly described by the eigenvalues and eigenvectors alone [15, section 5.7] [121, chapter 26]. (We stress, however, that there are many linear systems with nonnormal coefficient matrices for which eigenvalues and eigenvectors do describe the convergence behavior.)

Preconditioning is usually vital to reduce or remove this stagnation, and accelerate convergence. However, the non-normality of the coefficient matrix makes it difficult to know a priori what should be achieved by a preconditioner. Despite this, many successful preconditioners have been proposed. For mildly convective problems, these may be based on the symmetric part of the coefficient matrix, that is, the part associated with the negative Laplace operator [122,123]. For convection-dominated problems, however, it is necessary to incorporate the nonsymmetric part. For example, the negative Laplace preconditioner can be replaced by stationary iterative methods that split the coefficient matrix into its Hermitian and skew-Hermitian parts [124,125].

Even the (block) Gauss-Seidel iteration may make for an effective preconditioner if variables are ordered in the direction of the flow [126,127] [128, chapter 7], and reordering is at the heart of incomplete factorization [129] and splitting preconditioners [130,131] for the convection-diffusion problem. For more complicated flows, particularly recirculation problems, it is challenging to order the variables well, so many preconditioners are based on simplifying the convection-diffusion operator [132,133]. A completely different option is to employ a matrix-equation solver for separable-coefficient problems as a preconditioner [134,135].

Multigrid preconditioners can also be effective, although advection causes difficulties that are not present for the Poisson problem. Gauss-Seidel smoothers can work well, but again for best effect the variables should be ordered according to the flow direction, or multiple sweeps in different directions should be used [80,128]. More generally, careful ordering of unknowns can improve multigrid preconditioners [136, section 10.4.3] [137,138]. In geometric multigrid methods, it is also important to avoid oscillations resulting from unstable coarse grid discretizations, which require more smoothing steps to remove. Options for dealing with oscillations include semi-coarsening [139,140], matrix-dependent transfer operators [68, section 7.7.5] [80, section 5.4], upwinding [137,141], or rediscrctizing the coarse grid operator at each level using an appropriate stable discretization [142]. Robust multigrid methods for the Navier-Stokes equations also tackle the difficult problem of building a scalable preconditioner for a convection-diffusion operator [143,144]. AMG methods can make good preconditioners for convection-diffusion problems [145-147] by automatically identifying layers and/or semi-coarsening. On the other hand, it may take many levels to reach an acceptably small “coarse grid” problem [147].

3.2 | Systems of PDEs

Systems of PDEs can arise for a number of reasons, the most obvious cause being interactions between different components, for example in fluid-structure interactions or the magnetohydrodynamics equations. Such systems also naturally occur when there are interactions between different quantities, for example the pressure and velocity of a fluid in the Navier-Stokes equations. Alternatively, it may be more convenient to write a single high-order equation as a system of lower-order equations; a classical example is the biharmonic equation, which can be reformulated in terms of two second-order equations.² Systems may also arise when it is helpful to separate different components of the discretized problem, for example in finite element discretizations where it can be advantageous to group variables corresponding to different types of basis functions [149]. Alternatively, it may be advantageous to distinguish equations corresponding to boundary nodes, which may exhibit different behavior, or may be discretized differently [150].

3.2.1 | Navier-Stokes problems

We start with coefficient matrices arising from general PDE systems. As an example we consider the primitive variables formulation of the steady incompressible Navier-Stokes equations, which couple the velocity \mathbf{u} and pressure p of a fluid via the following equations:

$$\begin{aligned} -\nu \nabla^2 \mathbf{u} + \mathbf{u} \cdot \nabla \mathbf{u} + \nabla p &= \mathbf{f}, \\ -\nabla \cdot \mathbf{u} &= 0, \end{aligned} \tag{5}$$

²On the other hand, it may instead be advantageous to move from a system of PDEs to a single higher-order PDE when developing preconditioners [148].

subject to appropriate boundary conditions, on a bounded connected domain $\Omega \subset \mathbb{R}^d$, $d \in \{2, 3\}$ [128, chapter 8]. The parameter $\nu > 0$ here represents the kinematic viscosity.

Since the Navier-Stokes equations are nonlinear, linearization is required, with popular choices including Newton and Picard iteration. We consider Picard iteration here, which gives rise to a sequence of *Oseen problems*. Care is also needed when discretizing (5) to ensure that the resulting method is stable; this is related to the inf-sup Ladyzhenskaya-Babuška-Brezzi (LBB) conditions [128, chapter 3]. If a stable discretization is used, such as an LBB-stable finite element pair or the marker-and-cell finite difference scheme, then each Oseen problem is a *saddle point system* of the form

$$\underbrace{\begin{bmatrix} A & B^T \\ B & 0 \end{bmatrix}}_{\mathcal{A}} \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} = \begin{bmatrix} \mathbf{b} \\ \mathbf{c} \end{bmatrix}. \quad (6)$$

Here, $A \in \mathbb{R}^{n \times n}$ contains the diffusive and convective terms while $B \in \mathbb{R}^{m \times n}$, $m \leq n$, represents the discrete negative divergence. We assume for simplicity that A is invertible and that B is full rank. (For conditions on the invertibility of \mathcal{A} see [151, section 3.2].) Note that the (2, 2) block in (6) may be replaced by a negative semi-definite matrix to stabilize a discretization, and many of the preconditioners we discuss can be adapted to deal with this structure.

One widely used option is to precondition \mathcal{A} monolithically, for example via a multilevel method such as multigrid or domain decomposition [151, section 11]. Alternatively, given the saddle point form of the Oseen equations, a number of preconditioners described in Section 4 can also be applied to (6). Provided ν is large enough, preconditioners designed for the Stokes equations (cf Section 3.2.2) might also be suitable.

Yet another option is to exploit the block structure of \mathcal{A} when designing preconditioners. For problems in which different blocks are associated with different physical quantities, or different PDEs, block preconditioners are often known as *physics-based preconditioners*. These block preconditioners may take advantage of the specific problem structure by, for example, separating different spatial dimensions [152, 153]. Alternatively, many popular preconditioners are based on the following “ideal” choices that exploit only the saddle point structure [154]:

$$\mathcal{P}_u = \begin{bmatrix} A & B^T \\ 0 & \pm S \end{bmatrix} \quad \text{or} \quad \mathcal{P}_l = \begin{bmatrix} A & 0 \\ B & \pm S \end{bmatrix},$$

where $S = BA^{-1}B^T$ denotes the (negative) Schur complement. Related block diagonal preconditioners can also be applied, but we defer discussion of these to Section 3.2.2. Both \mathcal{P}_u and \mathcal{P}_l are ideal, since a method like preconditioned GMRES will converge in at most two iterations with either choice. On the other hand, the size of A makes forming S and applying \mathcal{P}_u or \mathcal{P}_l too costly in general, and so many practical preconditioners are based on approximating A and S [128, 155–157]. We stress that this block triangular approach can be applied to any saddle point problem, for example porous media flow [158, 159].

3.2.2 | Stokes problems

When advective inertial forces are small compared with viscous forces in (5) we can approximate the Navier-Stokes equations by the linear Stokes equations:

$$\begin{aligned} -\nabla^2 \mathbf{u} + \nabla p &= \mathbf{f}, \\ -\nabla \cdot \mathbf{u} &= 0, \end{aligned}$$

posed within $\Omega \subset \mathbb{R}^d$, $d \in \{2, 3\}$, with appropriate boundary conditions.

Discretization of these equations by, for example, an LBB-stable finite element pair or the marker-and-cell finite difference method will lead to a saddle point problem of the form (6) in which \mathcal{A} is symmetric. Specifically, $A \in \mathbb{R}^{n \times n}$ is a discrete representation of the negative vector Laplacian and $B \in \mathbb{R}^{m \times n}$, $m \leq n$, represents the discrete negative divergence (cf Section 3.2.1). We assume for simplicity that A is symmetric positive definite and that B has full rank although, as for the Navier-Stokes equations, these conditions may be relaxed [151].

The matrix \mathcal{A} is symmetric but indefinite, making it somewhat challenging to solve (6). It is possible to instead solve the symmetric positive (semi-)definite Schur complement system

$$S\mathbf{y} = BA^{-1}B^T\mathbf{y} = BA^{-1}\mathbf{b} - \mathbf{c}, \quad (7)$$

where $S = BA^{-1}B^T$ is again the (negative) Schur complement (cf Section 4), but since our focus here is on dealing with systems of PDEs we instead focus on solving the full system (6).

Given the similarity of the discretized Stokes equations with the Oseen problem, it is not surprising that many of the preconditioners discussed there can be adapted for the Stokes equations. However, since \mathcal{A} is symmetric, it may be desirable to use a symmetric Krylov method, such as preconditioned MINRES, which requires a symmetric positive definite preconditioner. Doing so precludes certain monolithic and block triangular preconditioners.

When a positive definite preconditioner is required, a natural choice for (6) is a block diagonal preconditioner, such as [154,160]

$$\mathcal{P} = \begin{bmatrix} A & 0 \\ 0 & S \end{bmatrix}, \quad (8)$$

for which preconditioned MINRES converges (in exact arithmetic) in at most three steps. On the other hand, as for the block triangular preconditioners for the Oseen problem, in practice A and S must be approximated. Since A is a vector Laplacian, methods for the Poisson equation (see Section 3.1.1) can be used in its approximation. Thus, as for many block problems, the challenge lies in approximating S . Luckily, analysis (see, eg, [128, chapter 4]) shows that when an LBB-stable finite element pair is used, a good approximation to S is the mass matrix on the pressure space, Q , or an approximation, for example its diagonal, a lumped version, or a fixed number of steps of Chebyshev semi-iteration [161-163]. Additionally,

$$\mathcal{P} = \begin{bmatrix} A & 0 \\ 0 & Q \end{bmatrix}$$

is an operator preconditioner [55]. For the marker-and-cell finite difference scheme, life is even easier since we can replace S by the identity matrix. We note that the preconditioner (8) has been generalized to certain problems containing more blocks [164-167], and that in general block diagonal preconditioning is a popular choice for symmetric saddle point problems [168-171]. The operator preconditioning approach can also be more widely employed; for more examples, see [55].

One aspect that is not often taken into account by block preconditioners is the appropriate scaling factors for the different blocks. Scaling can have an effect on the number of iterations needed for convergence [172,173]. More fundamentally, different equations are expressed in terms of different physical units, and these units may not be faithfully represented in the preconditioned system without scaling the blocks by relevant physical parameters, for example the viscosity in the Stokes equations [174]. This need becomes more obvious when dealing with varying parameters, as in two-phase flow [155,175].

3.3 | Toeplitz matrices

When constant-coefficient PDEs are discretized on uniform grids on simple domains, *Toeplitz* and *multilevel Toeplitz* matrices may arise. In this case, the highly structured nature of these matrices can be exploited to construct fast solvers. Although these structures have been utilized in the context of PDEs for some time [176,177], recent interest in the discretization of fractional differential equations (FDEs) has proved a new source of (multilevel) Toeplitz problems [178,179]. This is because nonlocal fractional operators make the solution of the required linear systems extremely challenging unless tailored approaches are employed.

When one-dimensional linear constant-coefficient ordinary differential equations with Dirichlet boundary conditions are discretized on a uniform grid by, for example, standard finite difference methods, the resulting coefficient matrix is

Toeplitz, that is, of the form

$$A = \begin{bmatrix} a_0 & a_{-1} & \dots & a_{-n+2} & a_{-n+1} \\ a_1 & a_0 & a_{-1} & & a_{-n+2} \\ \vdots & a_1 & a_0 & \ddots & \vdots \\ a_{n-2} & & \ddots & \ddots & a_{-1} \\ a_{n-1} & a_{n-2} & \dots & a_1 & a_0 \end{bmatrix}.$$

If the boundary conditions are instead periodic then A is not only Toeplitz, but also *circulant*, that is, $(A)_{ij} = (A)_{(i-j) \bmod n}$. In two or more spatial dimensions the discretization of PDEs no longer leads to Toeplitz matrices in general. However, when linear constant-coefficient problems are discretized on uniform tensor-product grids we may obtain multilevel Toeplitz matrices. The best known of these are block Toeplitz with Toeplitz block (BTTB) matrices for two-dimensional problems.

Although for ordinary or partial differential equations the resulting (multilevel) Toeplitz matrices are typically sparse, the nonlocality of fractional operators means that matrices arising from FDE applications are dense. It is unusual for Krylov subspace methods to be competitive with direct methods for dense matrices, but here we can compute matrix-vector products in $O(n \log(n))$ operations by embedding A in a $2n \times 2n$ circulant matrix and utilizing the fast Fourier transform (FFT) [180, section 4.8.2].

Toeplitz matrices from differential equations are usually generated by a symbol, f , in the sense that the matrix entries are its Fourier coefficients, that is,

$$a_k = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(\theta) e^{-ik\theta} d\theta.$$

(Multilevel Toeplitz matrices lead to multivariate symbols.) This connection with the symbol is extremely powerful, and many properties of A can be deduced from f . For example, if $f \in L^1((-\pi, \pi))$ is real-valued then A must be Hermitian and

$$\operatorname{ess\,inf}_{\theta \in (-\pi, \pi)} f(\theta) \leq \lambda_{\min}(A) \leq \lambda_{\max}(A) \leq \operatorname{ess\,sup}_{\theta \in (-\pi, \pi)} f(\theta),$$

where the inequalities are strict if f is not constant almost everywhere. Asymptotic (in n) eigenvalue or singular value distributions of A can also be determined from f [181-184]; since the symbols of differential and integral operators typically have zeros, these results clearly illustrate why A is often extremely ill-conditioned for large n (as is well known).

Noticing that the inverse of a circulant matrix can also be applied to a vector in $O(n \log(n))$ operations using the FFT, Strang [185] and Olkin [186] proposed certain circulant preconditioners. This led to the development of a number of other preconditioners based on circulants [187,188], fast transforms [189,190], banded Toeplitz matrices [191,192], and multigrid [193] (see the survey by Chan and Ng [194] and the books [195,196]). For multilevel Toeplitz matrices, multilevel circulant preconditioners are not optimal since iteration numbers increase with n [197,198] but multigrid methods, for example, can give mesh-independent convergence rates [199], and semi-circulant preconditioners can be effective for certain PDE problems [176,177].

The vast majority of preconditioners proposed for Toeplitz problems were originally designed for Hermitian matrices generated by a smooth symbol f , for which estimates of the eigenvalues of the preconditioned matrix can be obtained and related to the convergence rate of the Krylov method [191,200]. More recently these sorts of results have been extended to a wider class of problems, known as (block) generalized locally Toeplitz matrices [201,202]. For non-Hermitian matrices it is often possible to prove that the singular values and/or eigenvalues of the preconditioned matrix are clustered. Unfortunately, with the exception of LSQR or LSMR (or similar symmetrizations), it is not possible to link these results to convergence rates of Krylov subspace methods. An exception can be made in specific cases, for example [177], in which the condition number of the eigenvector matrix is also bounded. Another possibility is to use the specific Toeplitz structure to symmetrize, as described in Section 5.

3.4 | Time-dependent problems

Time-dependent problems present new, often significant, challenges for efficient numerical methods. The most common way to solve time-dependent PDEs is by a method-of-lines approach, which gives rise to a sequence of problems (one for each time step), that must be solved sequentially. In this case, preconditioners are typically based on those for the analogous steady problem. These time-dependent systems can sometimes be easier to solve than their steady counterparts, in the sense that matrix blocks can have more convenient spectral properties. This is because time-stepping typically introduces a positive definite matrix, for example a mass matrix or identity matrix scaled by the inverse of the time step. For large time steps, however, the effect is usually small, and a good preconditioner should work for the range of time steps appropriate for the given problem.

An alternative approach is to solve for all time steps simultaneously. This is sometimes called an “all-at-once” or “one-shot” method. Any parallel time integration scheme that can be described by a linear operator may be used as a preconditioner, with an overview of possible methods, for example domain decomposition and multigrid approaches, given in the survey by Gander [203]. Other techniques include matrix equation-based preconditioners [204] (with a flexible Krylov method if the matrix-equation solver is nonlinear) or, in the case of uniform time steps, preconditioners that exploit block Toeplitz structure [205,206]. A disadvantage of the all-at-once approach is that the sizes of all time steps need to be determined in advance. However, this can still be useful in certain settings, for example predictor-corrector methods [207] or in optimization problems [208-210].

3.5 | Problems with heterogeneous coefficients

Problems with heterogeneous coefficients are in general more difficult to precondition than those without. They occur frequently when modeling real-world phenomena, often arising in nonlinear problems. The simplest such problem is a diffusion problem with varying diffusivity, that is, (3) with nonconstant α . An elementary option that can be effective for such problems is to first scale the coefficient matrix by its diagonal [211], which can reduce the effect of α , and then apply a standard Poisson preconditioner. Alternatively, many of the techniques described above, such as multigrid [212,213] and domain decomposition [88,211,214], have been adapted to deal with problems with heterogeneous coefficients by, for example, using more sophisticated smoothers, scalings, or coarse grid solves. Recently, insights into the performance of simple Laplace preconditioners for (3) were obtained via careful eigenvalue analysis [67]; this, similarly to [211], showed that in the case of heterogeneous problems, looking at the condition number of the preconditioned coefficient matrix alone may not be indicative of the effectiveness of the preconditioner. Finally, we note that it may be advantageous to reformulate the system. For example, (3) can be rewritten as a system of first-order equations, for which preconditioners may be more readily constructed [168,215].

3.6 | Other methods

A *sparse approximate inverse (SPAI) preconditioner* for A may be defined as $P^{-1} = R$, where R minimizes $\|I - AR\|$ subject to R also possessing some sparsity pattern. This is most commonly considered in the Frobenius norm (as initially investigated in [216], see also [217-223]), for which the minimization problem then reduces to separate least-squares problems for each column of R that may be solved in parallel. Among many important works on the subject, we refer to [4, chapter 5] for a detailed summary of such preconditioners, to [224] for a description of Grote and Huckle’s successful SPAI algorithm for dynamically defining a sparsity pattern for R , to [225] for a discussion of nonsymmetric linear systems, and to [226-228] for comparative studies of different SPAI preconditioners and sparsity structures. In the context of PDEs, a key application is the use of Frobenius norm-based preconditioners for large dense systems resulting from boundary integral formulations of electromagnetic problems [217,218,229,230]. Other important applications of approximate inverse preconditioners arise in elasticity problems [219,222,223], magnetohydrodynamics [231], and acoustic simulations [220].

It is also possible to construct preconditioners based on fixed point iterations. Although simple approaches based on Jacobi and SOR are not often used on their own for PDE problems, splittings that are tailored to the PDE in question can form the basis of effective preconditioners. One example is a splitting according to the spatial dimension, which is the basis for the alternating directions implicit method [232] and related approaches like the dimension splitting method [152,153], and the Hermitian-skew-Hermitian splitting method (see Section 3.1.3).

A *deflation preconditioner* is a projection that aims to (approximately) remove a subspace responsible for slow convergence [19,86,233,234]. A common motivation is to remove the eigenspace associated with small eigenvalues, although in theory any subspace could be removed. On the other hand, it is also possible to precondition to augment the Krylov subspaces in which a solution is sought by adding search directions that accelerate convergence [19,234]. Closely related to these ideas are polynomial preconditioners, that is, preconditioners of the form $P^{-1} = p(A)$, for some polynomial p . These aim to incorporate spectral information, so that the properties of $P^{-1}A\mathbf{x} = P^{-1}\mathbf{b}$, where $P^{-1}A = AP^{-1} = Ap(A)$, are more favorable [16, section 12.3].

We have already seen how dense Toeplitz problems may be solved by Krylov subspace methods. Iterative methods can also be applied to other dense, but data-sparse, matrices stored in a format that enables fast matrix-vector products, where a matrix is data-sparse if it can be stored in some representation using far fewer than n^2 degrees of freedom. For PDEs, the most common structure is the presence of low-rank off-diagonal blocks. Intuitively, this is because far-range interactions tend to be relatively weak, and so can be well approximated by low-rank structures. Data-sparse formats include \mathcal{H} -matrices [136,235,236], \mathcal{H}^2 matrices [237], hierarchically semi-separable matrices [238], the fast multipole method [239], and recursive skeletonization [240]. Often, by storing matrices in these formats, it is possible to accelerate matrix-vector products within Krylov subspace methods. However, it may also be possible to use these approaches for preconditioning [241,242].

We conclude this section by commenting on a technique that can be applied when a problem has two (or more) preconditioners that accelerate the convergence rate of a linear system. While it may be difficult to formulate a new preconditioner that captures the good behavior of each individual preconditioner, they can be combined within a *multipreconditioned iterative solver* [243,244]; this has been successfully used in the solution of parameter-dependent problems [245] and two-phase flow [246]. More generally, when dealing with sequences of problems it can make sense to reuse information from previously computed systems, including preconditioners [19, section 14] (see also Section 4.4).

4 | PRECONDITIONERS FOR OPTIMIZATION PROBLEMS

Another major field of research which necessitates the solution of linear systems of equations describing a physical process is that of optimization. As an illustration, if one wishes to solve the constrained *quadratic programming* problem:

$$\begin{aligned} \min_{\mathbf{x}} \quad & \frac{1}{2} \mathbf{x}^T H \mathbf{x} - \mathbf{b}^T \mathbf{x} \\ \text{s.t.} \quad & B \mathbf{x} = \mathbf{c}, \end{aligned}$$

where H is symmetric positive semidefinite, one may construct the Lagrangian:

$$\mathcal{L}(\mathbf{x}, \mathbf{y}) = \frac{1}{2} \mathbf{x}^T H \mathbf{x} - \mathbf{b}^T \mathbf{x} + \mathbf{y}^T (B \mathbf{x} - \mathbf{c}).$$

Finding a stationary point of \mathcal{L} then leads to a linear system exactly of the saddle point form (6), with $A = H$. Further matrices may arise within this system: for example, when solving problems with additional inequality constraints imposed on \mathbf{x} using an *interior point method*, terms arising from a barrier function enforcing these constraints lead to an additional diagonal matrix which is added to A [247]. Further matrix terms can also arise in the (2,2)-block of the saddle point system as a result of regularization terms applied to solve optimization problems.

Often, for example in the case of linear programming problems [248-251], A can be a diagonal matrix as $H = 0$. In this case it is helpful to reduce the system to the equivalent Schur complement system (7) (often referred to as the normal equations within the optimization literature), provided A is invertible.

In optimization it is important to consider the advantages and disadvantages of solving the Schur complement system or the saddle point system, especially if A^{-1} has a particularly convenient structure. The Schur complement is guaranteed to be symmetric positive semidefinite; however if B has full row rank the reduced system may be significantly more ill-conditioned than the matrices A and B themselves, and moreover with increased density. Examining instead the saddle point system circumvents these issues, however this system is indefinite and is (often significantly) larger than the Schur complement matrix.

Another important consideration is that, in optimization, frequently very little is known about the structure of the linear system, apart from the matrix being sparse in many applications (which in general we assume in this section).

This contrasts with preconditioning PDE problems, where some more descriptive properties of the system are generally known a priori, due to the features of the analogous infinite-dimensional problems for instance. Therefore, the strategies used to precondition systems in optimization often need to be more general, to account for varying structures arising in the sub-blocks of the linear systems.

In this section we give a representative summary of some classes of methods used to precondition linear systems arising from optimization problems in the literature. We wish to emphasize that many strategies arising from the two areas of PDEs and optimization can have considerable similarities and can sometimes be transferred between the fields, notwithstanding the particular challenges that optimization problems pose. For instance, by exploiting the saddle point structures of matrices that often result, block preconditioners for PDEs like those in Section 3.2 can also be applied to linear systems arising from constrained optimization problems. It is also possible to apply many of the approaches outlined in this section to systems arising from PDEs, and we give examples in Section 4.6, which considers optimization problems with PDEs themselves acting as constraints. We present the following methods for optimization problems in this way because we wish to provide a self-contained summary of some preconditioners within the field of optimization, which we believe cannot be found in much of the literature, and which highlights in particular some methods that require relatively little knowledge about the precise structures of the matrices being tackled. In particular, some of the preconditioners we propose—notably incomplete factorizations—can be applied to *any* coefficient matrix A , not just the optimization problems considered in this section.

4.1 | Incomplete factorizations

As discussed previously, solving large-scale problems using direct methods can rapidly become impractical as the dimension of the system increases. Computing an LU factorization of a matrix $M = LU$ (or $M = LDU$) can result in L , U being significantly less sparse than M , a feature known as *fill-in*, although there are strategies in place to mitigate this including reorderings [1,252,253]. The same also applies to computing a Cholesky factorization $M = LL^T$ of a Hermitian positive definite matrix (or a factorization $M = LDL^T$ of systems that need not be positive definite). We note that M could denote a matrix relating to the entire system being solved, or a sub-block of it.

The idea of *incomplete factorization* is that one attempts to form an approximation of such a factorization, that at the same time is cheap to apply. As with preconditioning in general, this necessitates a trade-off between computational efficiency and accuracy of approximation, which in this case corresponds to how aggressively one may drop fill-in versus how potent an approximation results from this strategy. Essentially, algorithms are required that determine relevant criteria for the dropping of fill-in, based on position in the matrix, or value/magnitude of the entry. An incomplete Cholesky factorization of a matrix M then leads to

$$M = LL^T + E,$$

where E is the error induced by the dropped terms. The same applies to incomplete LU (ILU) factorization. There has been a substantial amount of research into devising incomplete LU and incomplete Cholesky factorizations, including within preconditioners [252,254-258].

One possible choice is to select sparsity patterns for the incomplete Cholesky or ILU factorizations which are the same as the matrix M itself. This leads to the ILU(0) or IC(0) factorizations. While these will be comparatively cheap to work with, they may contain rather little information about the matrix M as a result. However, this has nonetheless been shown to be effective for preconditioning certain classes of matrices [13,259]. This methodology can be generalized to IC(ℓ) or ILU(ℓ) preconditioners (where ℓ is a positive integer) to achieve different levels of fill-in. We refer to [4, sections 3 and 4] for a thorough survey of such methods, as well as research on the existence and stability of block factorizations and preconditioners.

We highlight that triangular factorizations need not be computed using Cholesky for positive definite matrices: for example in [260] Benzi and Tuma compute such a factorization by orthogonalization with respect to the inner product generated by M . Among much other research in the area of incomplete factorization preconditioners, there has also been a focus on multilevel and multifrontal preconditioners [257,258,261,262], some of which can give rise to parallel computing technologies.

Within optimization, incomplete factorizations form a valuable tool, which can be used as an effective preconditioner for the normal equations, or as an approximation for one block of a saddle point system. There have been many such

applications within the field, for example preconditioning linear systems arising from the solution of linear programming problems [248,263] (see also [249]), nonlinear optimization [264,265], least-squares problems (as in Section 4.5), and many others. Incomplete factorizations are also widely used in applications beyond optimization, since they can be computed using only matrix entries.

4.2 | Constraint preconditioning

One key class of preconditioners for saddle point systems arising in optimization, of the form (6), is that of *constraint preconditioners* [266-268]:

$$\mathcal{P} = \begin{bmatrix} \hat{A} & B^T \\ B & 0 \end{bmatrix}. \quad (9)$$

These preconditioners are so called because at each iteration of an appropriate Krylov method, the constraints (corresponding to the matrix B) are exactly satisfied in exact arithmetic [266,269]. If A is symmetric positive definite on the kernel of B , which for this discussion we assume is of full row rank, then \hat{A} is typically selected to be another symmetric positive definite matrix.

We highlight that \mathcal{P} is not a symmetric positive definite matrix, and yet it may be applied within Krylov methods such as CG and MINRES: this being equivalent to the *projected CG* and *projected MINRES* algorithms [269], as the preconditioner is positive definite on the constraint manifold [266]. Furthermore, all eigenvalues of the preconditioned linear system $\mathcal{P}^{-1}A$ that are not equal to one, are equal to those of $\hat{A}^{-1}A$ [267]. An effective constraint preconditioner requires a method to construct a matrix of the form (9), ideally such that the resulting matrix \hat{A} is a potent approximation of A and such that its inverse is cheap to apply to a vector.

One widely used approach for constructing such a preconditioner is to make use of a *Schilders factorization* [9,270,271], whereby one decomposes (possibly after reordering of the linear system) $B = [B_1 \ B_2]$ such that B_1 is invertible, then takes as a preconditioner:

$$\mathcal{P} = \begin{bmatrix} B_1^T & 0 & L_1 \\ B_2^T & L_2 & G \\ 0 & 0 & I \end{bmatrix} \begin{bmatrix} D_1 & 0 & I \\ 0 & D_2 & 0 \\ I & 0 & 0 \end{bmatrix} \begin{bmatrix} B_1 & B_2 & 0 \\ 0 & L_2^T & 0 \\ L_1^T & G^T & I \end{bmatrix}.$$

This factorization is guaranteed to equate to a matrix of the form (9), for any D_1, L_1, G of the correct dimensions, and for nonsingular matrices D_2, L_2 . The key task at this point is to construct matrices that well approximate the (1, 1)-block of the saddle point system (6). We note that alternative factorizations may also be derived, for instance based on null-space factorizations (see [272] for a summary of some such approaches, and [151] for a statement in the context of constraint preconditioning).

Constraint preconditioners may also be constructed for generalized saddle point systems, of the form (6) but with a nonzero (2, 2)-block. Such systems can arise from optimization problems with additional regularization, and associated constraint preconditioners are considered in, for example, [265,270,271,273].

Within the field of optimization, constraint preconditioners have been devised for linear and quadratic programming problems in [247,248,266,267,273-275], and additionally for nonlinear optimization problems in [265,267,268,271,276,277].

We highlight a number of further studies into the field of constraint preconditioning. Theoretical properties and convergence results are studied in detail in [278], constraint preconditioners for nonsymmetric indefinite linear systems are analyzed in [279], and a new limited memory approach with the objective of improving the stability and robustness of constraint preconditioning is described in [280]. Such preconditioners have also been devised for optimization problems with PDE constraints [281], including within the field of topology optimization involving PDEs [282].

4.3 | Augmented Lagrangian preconditioning

Another class of preconditioners for saddle point systems (6) from optimization are so-called *augmented Lagrangian preconditioners*:

$$\mathcal{P} = \begin{bmatrix} A + B^T W^{-1} B & 0 \\ 0 & W \end{bmatrix}. \quad (10)$$

This type of preconditioner is particularly useful if the $(1, 1)$ -block A of (6) is singular, but is positive definite on the kernel of B . In this case many classical saddle point preconditioners cannot be straightforwardly applied, but (10) is invertible for any symmetric positive definite matrix W , and furthermore is symmetric positive definite so may be applied within the MINRES algorithm. The eigenvalues of the resulting preconditioned system are analyzed in [283], demonstrating many enjoyable properties, and much original work on such preconditioners was tailored to solving PDEs, see [143, 144, 284] for instance. Of course, the effectiveness of the preconditioner depends on a suitable choice of W . One frequent choice is that of $W = \omega^{-1}I$, which is analyzed in detail in [285], especially in relation to the selection of the parameter $\omega > 0$.

One link with the field of optimization (relating in particular to the naming of the preconditioner) is through the augmented Lagrangian idea in optimization [9, 286, 287], particularly useful in cases where A is an ill-conditioned matrix, which notes that the system (6) is equivalent to the system:

$$\begin{bmatrix} A + B^T W^{-1} B & B^T \\ B & 0 \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} = \begin{bmatrix} \mathbf{b} + B^T W^{-1} \mathbf{c} \\ \mathbf{c} \end{bmatrix}.$$

Indeed the augmented Lagrangian preconditioner has been applied widely in optimization: one of the first such applications was to primal-dual interior point methods for linear and quadratic programming problems [288], see also [285, 289] for other applications in optimization. The augmented Lagrangian preconditioner idea has been extended to nonsymmetric saddle point systems in [143, 290], and has also found applicability to regularized systems where the $(2, 2)$ -block of the saddle point system is nonzero [289, 291]. It is also possible to construct block triangular analogues of (10) for saddle point systems [9, 143, 288, 290, 291].

4.4 | Low-rank updates of preconditioners for sequences of linear systems

A research area of active interest of late within the optimization community is the solution of sequences of linear systems

$$M^{(k)} \mathbf{x} = \mathbf{b}^{(k)}, \quad k = 0, 1, 2, \dots,$$

each of which may be of saddle point form (6), for instance. These can arise, for example, through the solution of nonlinear equations or an optimization problem using a Newton- or Broyden-type method, yielding a sequence of subproblems of the form

$$J(\mathbf{x}_k)(\mathbf{x}_{k+1} - \mathbf{x}_k) = -F(\mathbf{x}_k), \quad k = 0, 1, 2, \dots,$$

where $J(\mathbf{x}_k)$ is the Jacobian evaluated at the current iterate \mathbf{x}_k . It has been found that constructing *low-rank updates* of preconditioners may be highly effective. For instance, assuming one starts with an effective preconditioner \mathcal{P}_0 for $J_0 := J(\mathbf{x}_0)$, one may successively “correct” preconditioners at each iteration using rank-1 updates for \mathcal{P}_k :

$$\mathcal{P}_{k+1} = \mathcal{P}_k + \mathbf{u}\mathbf{v}^T, \quad k = 0, 1, 2, \dots,$$

and then use the Sherman-Morrison formula to apply \mathcal{P}_{k+1}^{-1} to $J_{k+1} := J(\mathbf{x}_{k+1})$ (see [292] for a discussion). This motivation may of course be generalized to updated preconditioners of higher ranks. In a similar spirit to that of SPAI preconditioners, for instance, such preconditioners may be constructed, with a reasonable objective being that $\|I - \mathcal{P}_k^{-1} J_k\|$ is bounded and small.

Crucial early work in this direction includes the papers of Martínez [293,294], where improving a preconditioner through low-rank updating is suggested. The preconditioners derived in these works were termed *secant preconditioners*, because they were required to satisfy the “secant property”:

$$P_{k+1}(\mathbf{x}_{k+1} - \mathbf{x}_k) = F(\mathbf{x}_{k+1}) - F(\mathbf{x}_k).$$

The idea of updating a previously computed matrix to accelerate the CG method, by saving information in the form of a quasi-Newton matrix, is also examined in [295], based on earlier research in this area such as in [296,297].

Preconditioners for the inexact Newton method were devised, for example, in [292,298], based on Broyden-type rank-1 updates [292] and BFGS rank-2 updates [298], respectively. Preconditioning of nonsymmetric systems using updates of incomplete factorizations is discussed in [299], and updated preconditioners which may be applied in a matrix-free setting are derived in [300].

Another research area of interest in optimization is the construction of low-rank updates of constraint preconditioners for quadratic programming problems [301-303]. In particular the paper [303] extends work in [304], where limited memory preconditioners for symmetric positive definite systems are considered. We also highlight the work [305], where low-rank updates of preconditioners are constructed for data assimilation, with application to the 4D-Var weather prediction model.

Finally, we note that low-rank updates of preconditioners can also be helpful when solving sequences of linear systems within applications other than optimization, for example the solution of PDEs (particularly time- or parameter-dependent problems), nonlinear problems, eigensolvers, and rational Krylov methods for matrix functions. In the particular, but important, case that these systems are all of the form $A = M + \psi I$, with ψ changing, modifying a given preconditioner for such a sequence is considered in, for example, [306-308]. We refer to [309] for a recent survey of low-rank updates of preconditioners, which considers a number of the above applications.

4.5 | Preconditioners for least-squares problems

Another specific class of problems that has gained much interest in recent years is that of *least-squares problems*:

$$\min_{\mathbf{x}} \|\mathbf{b} - A\mathbf{x}\|_2. \quad (11)$$

Differentiating the square of the functional in (11) with respect to \mathbf{x} to find the stationary point leads to the *normal equations*:

$$A^T A \mathbf{x} = A^T \mathbf{b}, \quad (12)$$

which in turn may easily be shown to be equivalent to solving the *augmented system*:

$$\begin{bmatrix} I & A \\ A^T & 0 \end{bmatrix} \begin{bmatrix} \mathbf{r} \\ \mathbf{x} \end{bmatrix} = \begin{bmatrix} \mathbf{b} \\ \mathbf{0} \end{bmatrix}, \quad \mathbf{r} = \mathbf{b} - A\mathbf{x}. \quad (13)$$

Similarly to when we consider the respective merits of solving saddle point systems of general form (6) versus Schur complement systems (7), there are advantages and disadvantages of solving the augmented system and the normal equations.

We first refer to two important articles summarizing developments in preconditioning for least-squares problems: [310] considers the solution of the normal equations using the conjugate gradient for least squares (CGLS) method before proposing new strategies, and [311] presents a thorough numerical study of preconditioners for the normal equations and the augmented system. We summarize the findings of these two papers very briefly here.

A number of highly effective strategies exist for preconditioning the normal equations (12). Apart from Jacobi (diagonal) preconditioning (see [311]), one may apply incomplete Cholesky methods (provided A has full column rank), including limited memory approaches [312-314]. One may also make use of incomplete orthogonalization methods [315,316], including incomplete QR (IQR) decompositions [317,318] based on Gram-Schmidt orthogonalization with

no dropping in Q , a method that is termed *compressed incomplete modified Gram-Schmidt*. IQR methods have also been extended to multilevel IQR approaches [319]. Another valuable class of approaches for the normal equations are robust and balanced incomplete factorizations, see [260,310,320].

Incomplete factorization strategies also exist for preconditioning the augmented system (13), such as signed incomplete Cholesky factorizations [321], or incomplete LDL^T factorizations [322].

We also highlight a number of other developments regarding the iterative solution of least-squares problems, including the limited memory preconditioner for the normal equations developed in [2], preconditioners based on approximate inverses [323,324], the solution of such problems based on augmented matrix methods [325, chapter 7], and preconditioners based on LU factorizations for both the normal equations [326] and the augmented system [327]. Specialized preconditioners can also be devised for least-squares problems with more specific properties, such as weighted Toeplitz least-squares problems [328] (cf Section 3.3).

4.6 | Link to PDEs and PDE-constrained optimization

As above, although PDEs and optimization form two separate branches of mathematics, there is considerable scope for many of the preconditioning strategies presented in this section to be used to tackle the linear systems discussed in Section 3 that arise from discretizations of PDEs. For example, ILU factorizations have been extensively applied to PDE problems (see [329,330] for instance), and incomplete factorizations find considerable applicability as smoothers within multigrid [68]. Constraint preconditioners [278,331-333] and augmented Lagrangian preconditioners [143,144,284,334] have also been widely used when solving PDEs.

However, it can also be that the nature of PDE and optimization problems are distinct, and so one must carefully tailor preconditioned iterative solvers to the problem at hand. For example, PDE problems typically offer more specific matrix structures which may be exploited by preconditioners, whereas optimization solvers frequently need to be more general. The differing philosophies between these two classes of problems is one reason why the methods presented in Sections 3 and 4 are distinct.

We conclude this section with one particular class of problems in which PDEs and optimization problems intersect: that of *PDE-constrained optimization*. To give an illustration of iterative solvers for the resulting linear systems, consider the *Poisson control* problem:

$$\begin{aligned} \min_{u,f} \quad & \frac{1}{2} \|u - \hat{u}\|_{L^2(\Omega)}^2 + \frac{\beta}{2} \|f\|_{L^2(\Omega)}^2 \\ \text{s.t.} \quad & -\nabla^2 u = f \text{ in } \Omega, \end{aligned} \tag{14}$$

given a domain $\Omega \subset \mathbb{R}^d$, $d \in \{2, 3\}$, a function \hat{u} , and a positive regularization parameter β , as well as appropriate boundary conditions. The variables u and f are often referred to as the *state variable* and *control variable*, respectively. Discretizing this problem using finite elements leads to a system of the form [52,335]:

$$\begin{bmatrix} Q & 0 & K \\ 0 & \beta Q & -Q \\ K & -Q & 0 \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \mathbf{y} \end{bmatrix} = \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \\ \mathbf{c} \end{bmatrix},$$

where Q denotes a mass matrix and K a stiffness matrix, and with $\mathbf{x}_1, \mathbf{x}_2$ corresponding to the discretizations of u, f .

Poisson control problems of this type have been considered from both the infinite-dimensional [52,55,167,174,336] and finite-dimensional [281,335,337,338] points of view. In particular by viewing the problem on the infinite-dimensional level, for instance by considering nonstandard norms, and then discretizing the resulting infinite-dimensional interpretation of the preconditioner, one may arrive at the optimal preconditioner [52,55]:

$$\mathcal{P} = \begin{bmatrix} \beta^{1/2}K + Q & 0 & 0 \\ 0 & \beta Q & 0 \\ 0 & 0 & \beta^{-1}(\beta^{1/2}K + Q) \end{bmatrix}.$$

This methodology may also be applied in the setting of constraint preconditioners [336], for instance. On the other hand, by considering the properties of the saddle point system purely on the discrete level, one may also apply the preconditioner [335]:

$$\mathcal{P} = \begin{bmatrix} Q & 0 & 0 \\ 0 & \beta Q & 0 \\ 0 & 0 & (K + \beta^{-1/2}Q)Q^{-1}(K + \beta^{-1/2}Q) \end{bmatrix},$$

which is also optimal. This approach may itself be extended to block triangular preconditioners [335,338], for instance. Both preconditioners lead to provably mesh- and β -independent solvers for the problem (14), and we note that both preconditioning “routes” lead to different (if related) structures.

Interestingly, both the “infinite-dimensional” and the “discrete” routes are known to deliver mesh- and parameter-robust solvers for particular PDE-constrained optimization problems, for which the other approach as yet cannot. For instance, deriving preconditioners on the infinite-dimensional level and then discretizing these operators has led to potent solvers for the Stokes control problem [52], and a number of time-periodic problems [339]. On the other hand, considering the optimization problem using a purely discrete (saddle point) framework has similarly generated robust solvers for convection-diffusion control problems [340], and distributed time-dependent problems [341].

We highlight that there also exist a range of other methods for the Poisson control problem (14), for example modified Hermitian and skew-Hermitian methods [342], nonsymmetric preconditioners which are not of block triangular form [343], and constraint preconditioners [281]. Of course adjustments can be made to problems of the form (14), that lead to more complex structures. Preconditioning linear systems arising from such problems is an active research area, with too many references to mention here. To give an illustration, there has been work on deriving preconditioners for problems with additional algebraic box constraints on the state and/or control variables [344–348], problems with limited observation [166], and parameter estimation problems [349], along with linear systems arising from a range of other scientific applications.

5 | PRECONDITIONERS WITH “NONSTANDARD” GOALS

In the previous two sections we have summarized a range of problems for which preconditioned iterative methods have found considerable utility, and a number of solution strategies that have been employed. In each case the key objective was to achieve rapid and robust convergence of an iterative method for solving a linear system $A\mathbf{x}=\mathbf{b}$, in a way which cannot be guaranteed by a direct method or an unpreconditioned iterative scheme. However this is not the limit of the capability of preconditioners. For instance one may also solve problems of a different general form using preconditioned solvers, such as eigenvalue problems, or alter the structure of an original problem statement. One may also have an additional motivation behind implementing a preconditioner in the first place: for instance reducing computer storage requirements, quantifying uncertainty in the solution, or implementing a method on high performance computing (HPC) architectures. In this section we very briefly survey valuable research which has been undertaken on preconditioning strategies for achieving such “nonstandard” objectives.

- **Eigenvalue problems:** Consider the problem of finding a simple interior eigenvalue λ of a symmetric matrix A , and its associated eigenvector \mathbf{x} . Algorithms for this include inverse iteration or Rayleigh quotient iteration (RQI), both of which require the solution of linear systems

$$(A - \sigma I)\mathbf{y}_k = \mathbf{x}_k, \quad (15)$$

where \mathbf{x}_k and $\mathbf{x}_{k+1} = \mathbf{y}_k/\|\mathbf{y}_k\|$ are successive approximations to an eigenvector \mathbf{x} .

When A is large, and matrix-vector products with A are easy to apply, it is attractive to solve (15) using a Krylov subspace method, such as MINRES. This leads to an inner-outer scheme, in which MINRES is the inner method, and inverse iteration or RQI forms the outer iteration. Although it is still possible to achieve the outer iteration convergence rate obtained when (15) is solved exactly, this generally necessitates that a more stringent MINRES tolerance is used as the outer iteration proceeds [350–354]. For a generic linear system, decreasing the tolerance would increase

the number of MINRES iterations, but perhaps surprisingly this is not the case here: as the outer iteration proceeds the right-hand side in (15) better approximates an eigenvector of $A - \sigma I$, making the system easier to solve. As a result, the number of inner iterations remains bounded [350,355]. (This clearly illustrates the role the right-hand side plays in determining the convergence rate of a Krylov subspace method.)

This property is usually destroyed when preconditioners are applied: although a good preconditioner will be effective at early outer iterations, decreasing the preconditioned MINRES tolerance typically increases the number of inner iterations. This is simply due to the fact that the preconditioned right-hand side no longer approximates an eigenvector of the preconditioned matrix. To remedy this, it is necessary to choose the preconditioner so that the preconditioned right-hand side approximates an eigenvector. This is a rather unorthodox view of preconditioning, but here proves to be just what is needed.

To see how this is achieved, consider the idealized situation that $(A - \sigma I)\mathbf{y}_k = \mathbf{x}$, that is, that the right-hand side is precisely the desired eigenvector. Then, for any positive definite preconditioner P , if $\mathbf{u} = (A - P)\mathbf{x}$,

$$P_{\text{tuned}} = P + \frac{\mathbf{u}\mathbf{u}^T}{\mathbf{x}^T\mathbf{u}}$$

is such that $P_{\text{tuned}}\mathbf{x} = A\mathbf{x}$. It can then be shown that preconditioned MINRES applied to this system would converge in exactly one step [355]. Although such a preconditioner is infeasible in general, replacing \mathbf{x} by \mathbf{x}_k and \mathbf{u} by $\mathbf{u}_k := (A - P)\mathbf{x}_k$ results in a preconditioner for which the number of inner iterations remains bounded as the outer iterations proceed. Moreover, by making use of the Sherman-Morrison formula, applying P_{tuned} is not significantly more challenging than applying P [180, section 2.1.4]. This idea has been analyzed for nonsymmetric generalized eigenvalue problems [356], and applied in inverse subspace iteration [357] and Arnoldi-type eigensolvers [358,359]. We have focused here on preconditioning the inner solves within iterative eigenvalue solvers. However, for some Krylov-based eigenvalue solvers, for example certain Lanczos-based methods, it is possible to precondition the eigenvalue solver itself to accelerate convergence to the desired eigenvalues [360-362].

- **Symmetrization:** As we saw in Section 2, MINRES is only applicable when a real coefficient matrix is symmetric, while CG additionally requires positive definiteness. It is somewhat surprising, therefore, that at least in theory any linear system with a real nonsymmetric coefficient matrix $A \in \mathbb{R}^{n \times n}$, can be solved using MINRES (or possibly CG) after appropriate preconditioning. This is because A can always be factored as $A = Y^{-1}T$ with Y, T symmetric [363], which implies that YA must be symmetric. Here, the preconditioner Y serves not to improve the convergence rate but to enable MINRES or CG to be used, with their short-term recurrences, optimality properties, and comprehensive convergence theory. All is not rosy, unfortunately, since computing Y usually requires the Jordan normal form of A , which is prohibitively expensive to compute, and almost always results in a dense matrix Y that is too costly to apply.

There are some exceptions, however: for certain structured matrices, a sparse Y is obtainable without recourse to the Jordan form. In these cases symmetrization is an attractive option. For example, if A is Toeplitz and

$$Y = \begin{bmatrix} & & 1 \\ & \ddots & \\ 1 & & \end{bmatrix},$$

then YA is a Hankel, hence symmetric, matrix. This well-known fact has been exploited to develop solvers that are typically faster than GMRES applied to the original system, thanks to the short-term recurrences of MINRES [364-366]. We stress that Y does not, in general, improve the convergence rate [367,368]; a second, symmetric positive definite, preconditioner is almost always required. However, in contrast to the nonsymmetric system, the preconditioner choice can be guided by MINRES convergence theory, and mathematical guarantees of fast convergence can be obtained. More generally, any persymmetric matrix is symmetrized by the reverse identity matrix, while $2n \times 2n$ Hamiltonian matrices are symmetrized by $\begin{bmatrix} 0 & I_n \\ -I_n & 0 \end{bmatrix}$.

A matrix that is symmetrized by Y as described above is self-adjoint with respect to the symmetric bilinear form [369, chapter 1] $\langle \mathbf{x}, \mathbf{y} \rangle_Y = \mathbf{y}^T Y \mathbf{x}$. When this bilinear form is an inner product it can be used to define Krylov subspace methods with optimality properties, such as Y -CG or Y -MINRES [19, section 13]. Even for standard CG and MINRES the choice of preconditioner essentially fixes the inner product in which the method is defined [370]. By

utilizing a different inner product, it may be possible to find a preconditioner P such that $P^{-1}A$ is self-adjoint in the Y -inner product, even if A and/or P are not symmetric [371,372].

Combinations of preconditioner and inner product need not only result in symmetry to be advantageous. For example, when the symmetric part of A , namely $Y = (A + A^T)/2$, is positive definite, $Y^{-1}A$ is the sum of the identity matrix and a matrix that is skew-adjoint with respect to the Y -inner product. In this case a short-term recurrence method can be applied [373,374].

- **Ill-posed problems:** For ill-posed problems $A\mathbf{x}=\mathbf{b}$, the matrix A typically has singular values that decay to zero and a right-hand side that is affected by noise, that is, $\mathbf{b} = A\mathbf{x}_{\text{true}} + \boldsymbol{\eta}$. For such problems, we would like to approximate \mathbf{x}_{true} . However, the ill-conditioning of A means that approximations to \mathbf{x} and \mathbf{x}_{true} may differ greatly even when $\|\boldsymbol{\eta}\|$ is small. Under certain conditions, iterative methods applied to ill-posed problems *do* approximate \mathbf{x}_{true} at early stages, but thereafter this noise dominates, at which point approximations to \mathbf{x}_{true} get worse again. When this phenomenon, which is known as semi-convergence [375-377], occurs the iterative method itself acts a regularizer of the problem.

A good preconditioner P will also make it easier to approximate \mathbf{x}_{true} , but this does not mean that we should choose $P \approx A$. Indeed, if P is a good representation of A it will almost certainly be ill-conditioned, and so its application may significantly amplify errors at early iterations. Instead, P should also regularize the problem somehow. Typically, one wants to damp the high-frequency components associated with small eigenvalues or singular values of A . A common option is to set these high-frequency components to one [378-380], but another option is to remove them completely via a singular preconditioner [381,382] or a projection [383]. We note that other types of regularization, for example Tikhonov, can also be enforced via preconditioning [384,385]. A different interpretation is to develop preconditioners for ill-posed problems using a Bayesian framework [386] (see the discussion on probabilistic solvers below).

- **Randomized preconditioning:** An area of significant recent interest in the linear algebra community is that of randomized linear algebra. In particular, randomized preconditioners may be developed for least-squares problems of the form (11). A key work is that of [387], where overdetermined problems are considered (that is, where $A \in \mathbb{C}^{m \times n}$, $m \geq n$), although the authors' methods can be extended to underdetermined regression. Here the problem (11) is replaced by a problem of minimizing $\|AP^{-1}\mathbf{y}-\mathbf{b}\|$ over vectors \mathbf{y} , whereupon the solution $\mathbf{x} = P^{-1}\mathbf{y}$. The solver incorporates a randomized transformation, with the preconditioner based on a permutation of the R factor from the QR decomposition of the transformed matrix A . Another important work is [388], where the randomized methods may be viewed as preconditioning the matrix A and vector \mathbf{b} with a data-independent random matrix P . In [389], a high-precision preconditioned solver is presented, where the rows and columns of A are blended, at which point a random sampling of rows is taken. We also refer to least-squares solvers based on random normal projection in [390,391], with [391] deriving a parallel solver for strongly over- or underdetermined systems, and randomized preconditioners for overdetermined ℓ_p regression problems (which involve minimizing $\|\mathbf{b}-A\mathbf{x}\|_p$ for $1 \leq p \leq \infty$) applied within Stochastic Gradient Descent-like solvers in [392].
- **Probabilistic linear solvers:** A further concept of considerable recent interest has been a Bayesian interpretation of preconditioning. A key early work in this direction is [393], where preconditioners for ill-posed systems of the form $A\mathbf{x}=\mathbf{b}$ (where A is of ill-defined rank) are obtained from the covariance matrix of the solution interpreted as a random variable. These were termed *priorconditioners*. Building on this, a statistical interpretation of left- and right-preconditioners was provided in [386], specifically connecting such preconditioners to covariance matrices of noise and prior, respectively, in certain settings. A priorconditioned CGLS method is applied to inverse problems from electrical impedance tomography in [394].

We also highlight research in [395,396] which provides novel derivations of CG methods, outputting probability measures which may be used to quantify uncertainty in the solution. In particular [395] extends probabilistic interpretations of secant methods and mentions the potential for preconditioning the derived method, and [396] discusses in detail preconditioners for the authors' *BayesCG* method, which are thought of as priors. As outlined in [397], the conceptual difference is that in [395] a posterior is constructed on A^{-1} whereas in [396] a posterior is constructed on the solution \mathbf{x} . The work [397] unifies the two approaches presented in [395,396], provides an interpretation of left- and right-preconditioning, and extends the idea to the GMRES algorithm.

- **Preconditioning and tensor methods:** An important consideration when solving linear systems, for instance those arising from high-dimensional PDEs, is that of data storage on a computer, which can become prohibitive unless specialized schemes are devised in order to combat this. One area of interest is that of utilizing low-rank tensor decompositions alongside preconditioned solvers, for linear systems which possess some underlying tensor-product structure. Such techniques assume in general that the input data and solution vector are approximable within such a low-rank tensor format, for instance the *hierarchical Tucker format* (see [398,399]), the *tensor train format* (see [400]),

and its extension to the *quantized tensor train format* (see [401,402]). Additionally, for linear systems with Kronecker product structure, for example arising from certain high-dimensional PDEs, it is also possible to use Krylov subspace methods designed for tensors or matrix equations [403,404]. Among many valuable works on tensor methods for the numerical solution of differential equations, coupled with preconditioners, we refer to [405-408] (see also [403] for the construction of Krylov subspace methods). Here, low-rank variants of iterative schemes are derived, where the iterates are replaced by low-rank tensors, and the potency of these schemes depends on the derivation of effective preconditioners. Additionally, recent research has been undertaken on tensor-based schemes for PDE-constrained optimization problems [409,410], as well as FDE-constrained optimization problems [411,412].

- **High-performance computing:** Some problems are so large and complex that they necessitate HPC and heterogeneous architectures. This brings new challenges for iterative methods: memory is usually limited and communication, particularly between nodes and processors, is slow relative to the speed of floating point operations. Since sparse matrix-vector products and vector operations are typically communication-bound, that is, their speed is limited by the speed of moving data, it is not straightforward to improve the efficiency of Krylov methods in HPC environments [413-415]. On the other hand, since Krylov methods rely on a few simple operations, there are opportunities to utilize, for example, graphics processing units (GPUs) to improve efficiency [416].

Although preconditioned iterative methods for HPC applications should also aim to solve the linear system as quickly as possible, rather than reducing the number of floating point operations the aim is typically to exploit parallelism and reduce communication costs. These goals can be achieved by improving data locality, thereby enabling more operations to be performed on data before communication is required, and/or by asynchronous communication and careful task scheduling, for example by using directed acyclic graphs to map data dependencies [417]. Unfortunately, some of the best preconditioners for serial computations, for example multigrid, are naturally communication bound as a result of coarse grid computations and grid transfer operations. Consequently, much work has gone into making these methods more suitable for HPC [418,419]. On the other hand, other methods including domain decomposition [89,90,92-95] and sparse approximate inverse preconditioners [420,421] have long been associated with HPC because they traditionally require relatively little communication and are highly parallelizable. However, perhaps counterintuitively, newer versions of these algorithms may increase communication and/or reduce parallelism to improve overall efficiency on cutting-edge computer architectures. For example, domain decomposition methods may also include a multilevel component that induces global coupling [422], while SPAI preconditioners may, for example, increase sequential computation to improve efficiency [423]. Additionally, it may be advantageous to tailor kernels to GPUs (see, eg, [424]).

Though the construction and application of incomplete factorization preconditioners can be problematic on parallel architectures, performance can be improved by, for example, using an asynchronous iterative method to form the factors [425] and exploiting techniques from sparse approximate inverses [426]. Other preconditioners have been developed with extreme-scale computing in mind; these may require more computation than alternatives, and hence are slower for serial or moderately parallel computations, but are better able to utilize many-core architectures [242,427].

6 | CONCLUSIONS

In this article we have given a short overview of the multitude of applications of preconditioning within scientific fields. The key goal of preconditioned iterative solvers is to ensure the rapid and robust solution of linear systems of equations, particularly in cases where this cannot be achieved by a direct method or unpreconditioned iterative solver, however in a number of cases one may have additional objectives in mind, as in Section 5. Research has ranged from the design of preconditioners for matrices about which a number of specific properties are known, particularly in PDE applications (as in Section 3), or more general black-box solvers, especially in the study of optimization problems (as in Section 4). Of course, due to the aim of providing a brief, overarching view on the vast field of preconditioning, we have omitted discussions on a number of other application areas, for example network problems and new developments in data science.

There does not yet exist a “universal” preconditioner, indeed well-constructed preconditioners tend to capture the properties arising from the particular application being considered, sometimes including the physics of the process itself. By employing this philosophy, many researchers have made significant contributions to the study of scientific processes through tackling the resulting linear systems.

ACKNOWLEDGEMENTS

The authors would like to thank three anonymous referees for their valuable comments and helpful suggestions. The authors would also like to express their gratitude to Luca Bergamaschi, Victorita Dolean, Jacek Gondzio, Zdeněk Strakoš, Jemima Tabcart, Ioana Vaduva, and Andy Wathen for their useful feedback on a draft of this work. John W. Pearson gratefully acknowledges support from the Engineering and Physical Research Council (EPSRC) grant EP/S027785/1, and a Fellowship from The Alan Turing Institute. Jennifer Pestana gratefully acknowledges support from the EPSRC grant EP/R009821/1. No new data were created for this publication.

REFERENCES

- [1] I. S. Duff, A. M. Erisman, and J. K. Reid, *Direct methods for sparse matrices*, 2nd ed., Oxford University Press, Oxford, UK, 2017.
- [2] S. Bellavia, J. Gondzio, and B. Morini, A matrix-free preconditioner for sparse symmetric positive definite systems and least-squares problems, *SIAM J. Sci. Comput.* **35** (2013), A192–A211.
- [3] A. V. Knyazev, Toward the optimal preconditioned eigensolver: Locally optimal block preconditioned conjugate gradient method, *SIAM J. Sci. Comput.* **23** (2001), 517–541.
- [4] M. Benzi, Preconditioning techniques for large linear systems: A survey, *J. Comput. Phys.* **182** (2002), 418–477.
- [5] A. J. Wathen, Preconditioning, *Acta Numer.* **24** (2015), 329–376.
- [6] D. Bertaccini and F. Durastante, *Iterative methods and preconditioning for large and sparse linear systems with applications*, CRC Press, Boca Raton, FL, 2018.
- [7] K. Chen, *Matrix preconditioning techniques and applications*, Cambridge University Press, Cambridge, MA, 2005.
- [8] O. Axelsson, A survey of preconditioned iterative methods for linear systems of algebraic equations, *BIT Numer. Math.* **25** (1985), 166–187.
- [9] M. Benzi and A. J. Wathen, *Some preconditioning techniques for saddle point problems*, in *Model Order Reduction: Theory, Research Aspects and Applications. Mathematics in Industry (The European Consortium for Mathematics in Industry)*, Vol **13**, W. H. A. Schilders, H. A. van der Vorst, and J. Rommes, Eds., Springer-Verlag, Berlin/Heidelberg, Germany, 2008, 195–211.
- [10] M. A. Olshanskii and E. E. Tyrtshnikov, *Iterative methods for linear systems: theory and applications*, Society for Industrial and Applied Mathematics, Philadelphia, PA, 2014.
- [11] G. Mele et al., *Preconditioning for linear systems*, Kindle Direct Publishing, 2020.
- [12] A. M. Turing, Rounding-off errors in matrix processes, *Q. J. Mech. Appl. Math.* **1** (1948), 287–308.
- [13] J. A. Meijerink and H. A. van der Vorst, An iterative solution method for linear systems of which the coefficient matrix is a symmetric M -matrix, *Math. Comput.* **31** (1977), 148–162.
- [14] A. Greenbaum, *Iterative methods for solving linear systems*, Society for Industrial and Applied Mathematics, Philadelphia, PA, 1997.
- [15] J. Liesen and Z. Strakoš, *Krylov subspace methods: Principles and analysis*, Oxford University Press, Oxford, UK, 2013.
- [16] Y. Saad, *Iterative methods for sparse linear systems*, Society for Industrial and Applied Mathematics, Philadelphia, PA, 2003.
- [17] M. Eiermann and O. G. Ernst, Geometric aspects of the theory of Krylov subspace methods, *Acta Numer.* **10** (2001), 251–312.
- [18] J. Liesen and P. Tichý, Convergence analysis of Krylov subspace methods, *GAMM-Mitt.* **27** (2004), 153–173.
- [19] V. Simoncini and D. B. Szyld, Recent computational developments in Krylov subspace methods for linear systems, *Numer. Linear Algebra Appl.* **14** (2007), 1–59.
- [20] A. N. Krylov, On the numerical solution of the equation by which the frequency of small oscillations is determined in technical problems, *Izv. Akad. Nauk SSSR Ser. Fiz.-Mat.* **4** (1931), 491–539.
- [21] M. R. Hestenes and E. Stiefel, Methods of conjugate gradients for solving linear systems, *J. Res. Nat. Bur. Stand.* **49** (1952), 409–436.
- [22] C. Lanczos, An iteration method for the solution of the eigenvalue problem of linear differential and integral operators, *J. Res. Nat. Bur. Stand.* **45** (1950), 255–282.
- [23] C. Lanczos, Solution of systems of linear equations by minimized iterations, *J. Res. Nat. Bur. Stand.* **49** (1952), 33–53.
- [24] J. K. Reid, *On the method of conjugate gradients for the solution of large sparse systems of linear equations*, in *Large Sparse Sets of Linear Equations*, J. K. Reid, Ed., Academic Press, London and New York, 1971, 231–254.
- [25] M. Engeli et al., *Refined iterative methods for computation of the solution and the eigenvalues of self-adjoint boundary value problems*, Birkhäuser, Basel, 1959.
- [26] C. C. Paige and M. A. Saunders, Solution of sparse indefinite systems of linear equations, *SIAM J. Numer. Anal.* **12** (1975), 617–629.
- [27] Y. Saad and M. H. Schultz, GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems, *SIAM J. Sci. Stat. Comput.* **7** (1986), 856–869.
- [28] Y. Saad, Krylov subspace methods for solving large unsymmetric linear systems, *Math. Comput.* **37** (1981), 105–126.
- [29] C. C. Paige and M. A. Saunders, LSQR: An algorithm for sparse linear equations and sparse least squares, *ACM Trans. Math. Softw.* **8** (1982), 43–71.
- [30] D. C.-L. Fong and M. Saunders, LSMR: An iterative algorithm for sparse least-squares problems, *SIAM J. Sci. Comput.* **33** (2011), 2950–2971.
- [31] R. W. Freund and N. M. Nachtigal, QMR: A quasi-minimal residual method for non-Hermitian linear systems, *Numer. Math.* **60** (1991), 315–339.

- [32] R. Fletcher, *Conjugate gradient methods for indefinite systems. Lecture notes in mathematics*, Vol **506**, Springer-Verlag, Berlin, Germany, 1976, 73–89.
- [33] H. A. van der Vorst, Bi-CGSTAB: A fast and smoothly converging variant of Bi-CG for the solution of nonsymmetric linear systems, *SIAM J. Sci. Stat. Comput.* **13** (1992), 631–644.
- [34] P. Wesseling and P. Sonneveld, *Numerical experiments with a multiple grid and a preconditioned Lanczos type method. Lecture notes in mathematics*, Vol **771**, Springer-Verlag, Berlin, Germany, 1980, 543–562.
- [35] P. Sonneveld and M. B. van Gijzen, IDR(s): A family of simple and fast algorithms for solving large nonsymmetric systems of linear equations, *SIAM J. Sci. Comput.* **31** (2008), 1035–1062.
- [36] M. Embree, *How descriptive are GMRES convergence bounds? Technical report NA-99-08*, University of Oxford, Oxford, UK, 1999.
- [37] A. Greenbaum and Z. Strakoš, *Matrices that generate the same Krylov residual spaces*, in *Recent Advances in Iterative Methods*, G. Golub, M. Luskin, and A. Greenbaum, Eds., Springer, New York, 1994, 95–118.
- [38] M. Arioli, V. Pták, and Z. Strakoš, Krylov sequences of maximal length and convergence of GMRES, *BIT Numer. Math.* **38** (1998), 636–643.
- [39] A. Greenbaum, V. Pták, and Z. Strakoš, Any nonincreasing convergence curve is possible for GMRES, *SIAM J. Matrix Anal. Appl.* **17** (1996), 465–469.
- [40] G. Meurant and J. Duintjer Tebbens, The role eigenvalues play in forming GMRES residual norms with non-normal matrices, *Numer. Alg.* **68** (2015), 143–165.
- [41] D. Braess and P. Peisker, On the numerical solution of the biharmonic equation and the role of squaring matrices for preconditioning, *IMA J. Numer. Anal.* **6** (1986), 393–404.
- [42] M. Arioli et al., Interplay between discretization and algebraic computation in adaptive numerical solution of elliptic PDE problems, *GAMM-Mitt.* **36** (2013), 102–129.
- [43] J. Papež, J. Liesen, and Z. Strakoš, Distribution of the discretization and algebraic error in numerical solution of partial differential equations, *Linear Algebra Appl.* **449** (2014), 89–114.
- [44] M. Arioli, E. H. Georgoulis, and D. Loghin, Stopping criteria for adaptive finite element solvers, *SIAM J. Sci. Comput.* **35** (2013), A1537–A1559.
- [45] M. Arioli, D. Loghin, and A. J. Wathen, Stopping criteria for iterations in finite element methods, *Numer. Math.* **99** (2005), 381–410.
- [46] P. Jiránek, Z. Strakoš, and M. Vohralík, A posteriori error estimates including algebraic error and stopping criteria for iterative solvers, *SIAM J. Sci. Comput.* **32** (2010), 1567–1590.
- [47] D. N. Arnold, R. S. Falk, and R. Winther, Preconditioning in $\mathbf{H}(\text{div})$ and applications, *Math. Comput.* **66** (1997), 957–984.
- [48] V. Faber, T. A. Manteuffel, and S. V. Parter, On the theory of equivalent operators and application to the numerical solution of uniformly elliptic partial differential equations, *Adv. Appl. Math.* **11** (1990), 109–163.
- [49] A. Klawonn, *Preconditioners for indefinite problems*. PhD Thesis, Universität Münster, 1996.
- [50] P. Oswald, On function spaces related to finite element approximation theory, *Z. Anal. Anwendungen* **9** (1990), 43–64.
- [51] P. Oswald, *Norm equivalencies and multilevel Schwarz preconditioning for variational problems. Forschungsergebnisse Math/92/01*, Friedrich Schiller Universität, Jena, Germany, 1992.
- [52] W. Zulehner, Nonstandard norms and robust estimates for saddle point problems, *SIAM J. Matrix Anal. Appl.* **32** (2011), 536–560.
- [53] U. Rüde, *Mathematical and computational techniques for multilevel adaptive methods*, Society for Industrial and Applied Mathematics, Philadelphia, PA, 1993.
- [54] J. Málek and Z. Strakoš, *Preconditioning and the conjugate gradient method in the context of solving PDEs*, Society for Industrial and Applied Mathematics, Philadelphia, PA, 2015.
- [55] K.-A. Mardal and R. Winther, Preconditioning discretizations of systems of partial differential equations, *Numer. Linear Algebra Appl.* **18** (2011), 1–40.
- [56] J. Hrnčíř, I. Pultarová, and Z. Strakoš, Decomposition into subspaces preconditioning: Abstract framework, *Numer. Alg.* **83** (2020), 57–98.
- [57] H. Yserentant, Hierarchical bases of finite-element spaces in the discretization of nonsymmetric elliptic boundary value problems, *Computing* **35** (1985), 39–49.
- [58] H. Yserentant, On the multi-level splitting of finite element spaces, *Numer. Math.* **49** (1986), 379–412.
- [59] J. Xu, Iterative methods by space decomposition and subspace correction, *SIAM Rev.* **34** (1992), 581–613.
- [60] L. Ling and E. J. Kansa, A least-squares preconditioner for radial basis functions collocation methods, *Adv. Comput. Math.* **23** (2005), 31–54.
- [61] D. Brown et al., On approximate cardinal preconditioning methods for solving PDEs with radial basis functions, *Eng. Anal. Bound. Elem.* **29** (2005), 343–353.
- [62] R. K. Beatson, J. B. Cherrie, and C. T. Mouat, Fast fitting of radial basis functions: Methods based on preconditioned GMRES iteration, *Adv. Comput. Math.* **11** (1999), 253–270.
- [63] R. E. Bank, Marching algorithms for elliptic boundary value problems. II: The variable coefficient case, *SIAM J. Numer. Anal.* **14** (1977), 950–970.
- [64] B. L. Buzbee, G. H. Golub, and C. W. Nielson, On direct methods for solving Poisson’s equations, *SIAM J. Numer. Anal.* **7** (1970), 627–656.
- [65] R. W. Hockney, A fast direct solution of Poisson’s equation using Fourier analysis, *J. ACM* **12** (1965), 95–113.
- [66] D. Fischer et al., On Fourier–Toeplitz methods for separable elliptic problems, *Math. Comput.* **28** (1974), 349–368.

- [67] T. Gergelits et al., Laplacian preconditioning of elliptic PDEs: Localization of the eigenvalues of the discretized operator, *SIAM J. Numer. Anal.* **57** (2019), 1369–1394.
- [68] U. Trottenberg, C. Oosterlee, and A. Schüller, *Multigrid*, Academic Press, London, 2001.
- [69] R. P. Fedorenko, A relaxation method for solving elliptic difference equations, *USSR Comput. Math. Math. Phys.* **1** (1962), 1092–1096.
- [70] R. P. Fedorenko, The speed of convergence of one iterative process, *USSR Comput. Math. Math. Phys.* **4** (1964), 227–235.
- [71] A. Brandt, Multi-level adaptive solutions to boundary-value problems, *Math. Comput.* **31** (1977), 333–390.
- [72] A. Brandt, S. McCormick, and J. Ruge, *Algebraic multigrid (AMG) for automatic multigrid solutions with application to geodetic computations. Technical report*, Institute for Computational Studies, Fort Collins, CO, 1982.
- [73] A. Brandt, S. McCormick, and J. Ruge, *Algebraic multigrid (AMG) for sparse matrix equations*, in *Sparsity and Its Applications*, D. J. Evans, Ed., Cambridge University Press, Cambridge, MA, 1985, 257–284.
- [74] A. Brandt, Algebraic multigrid theory: The symmetric case, *Appl. Math. Comput.* **19** (1986), 23–56.
- [75] J. E. Dendy, Black box multigrid, *J. Comput. Phys.* **48** (1982), 366–386.
- [76] J. W. Ruge and K. Stüben, *Algebraic multigrid*, in *Multigrid Methods*, S. F. McCormick, Ed., Society for Industrial and Applied Mathematics, Philadelphia, PA, 1987, 73–130.
- [77] P. Wesseling and C. W. Oosterlee, Geometric multigrid with applications to computational fluid dynamics, *J. Comput. Appl. Math.* **128** (2001), 311–334.
- [78] J. Xu and L. Zikatanov, Algebraic multigrid methods, *Acta Numer.* **26** (2017), 591–721.
- [79] W. Hackbusch, *Multi-Grid methods and applications*, Springer-Verlag, Berlin/Heidelberg, Germany, 1985.
- [80] P. Wesseling, *An introduction to multigrid methods*, John Wiley & Sons, New York, NY, 1992.
- [81] H. A. Schwarz, Über einen Grenzübergang durch alternierendes Verfahren, *Vierteljahrsschr. Naturf. Ges. Zürich.* **15** (1870), 272–286.
- [82] M. Dryja and O. Widlund, *An additive variant of the Schwarz alternating method for the case of many subregions. Technical report 339*, Department of Computer Science, New York University, New York, 1987.
- [83] M. Dryja and O. B. Widlund, *Chapter 16 – Some domain decomposition algorithms for elliptic problems*, in *Iterative Methods for Large Linear Systems*, D. R. Kincaid and L. J. Hayes, Eds., Academic Press, San Diego, 1990, 273–291.
- [84] M. Benzi et al., Algebraic theory of multiplicative Schwarz methods, *Numer. Math.* **89** (2001), 605–639.
- [85] A. Frommer and D. B. Szyld, An algebraic convergence theory for restricted additive Schwarz methods using weighted max norms, *SIAM J. Numer. Anal.* **39** (2001), 463–479.
- [86] R. A. Nicolaides, Deflation of conjugate gradients with applications to boundary value problems, *SIAM J. Numer. Anal.* **24** (1987), 355–365.
- [87] F. Nataf et al., A coarse space construction based on local Dirichlet-to-Neumann maps, *SIAM J. Sci. Comput.* **33** (2011), 1623–1642.
- [88] N. Spillane et al., Abstract robust coarse spaces for systems of PDEs via generalized eigenproblems in the overlaps, *Numer. Math.* **126** (2014), 741–770.
- [89] V. Dolean, P. Jolivet, and F. Nataf, *An introduction to domain decomposition methods: Algorithms, theory, and parallel implementation*, Society for Industrial and Applied Mathematics, Philadelphia, PA, 2015.
- [90] M. J. Gander, Schwarz methods over the course of time, *Electron. Trans. Numer. Anal.* **31** (2008), 228–255.
- [91] W. Hackbusch, *Iterative solution of large sparse systems of equations*, Springer, New York, NY, 1994.
- [92] T. P. A. Mathew, *Domain decomposition methods for the numerical solution of partial differential equations*, Springer-Verlag, Berlin/Heidelberg, Germany, 2008.
- [93] A. Quarteroni and A. Valli, *Domain decomposition methods for partial differential equations*, Oxford University Press, Oxford, UK, 1999.
- [94] B. Smith, P. Bjørstad, and W. Gropp, *Domain decomposition: Parallel multilevel methods for elliptic partial differential equations*, Cambridge University Press, Cambridge, MA, 1996.
- [95] A. Toselli and O. Widlund, *Domain decomposition methods – algorithms and theory*, Springer-Verlag, Berlin/Heidelberg, Germany, 2005.
- [96] R. E. Bank, A comparison of two multilevel iterative methods for nonsymmetric and indefinite elliptic finite element equations, *SIAM J. Numer. Anal.* **18** (1981), 724–743.
- [97] H. Calandra, S. Gratton, and X. Vasseur, *A geometric multigrid preconditioner for the solution of the Helmholtz equation in three-dimensional heterogeneous media on massively parallel computers*, in *Modern Solvers for Helmholtz Problems*, D. Lahaye, J. Tang, and K. Vuik, Eds., Birkhäuser, Cham, 2017, 141–155.
- [98] H. C. Elman, O. G. Ernst, and D. P. O’Leary, A multigrid method enhanced by Krylov subspace iteration for discrete Helmholtz equations, *SIAM J. Sci. Comput.* **23** (2001), 1291–1315.
- [99] E. Haber and S. MacLachlan, A fast method for the solution of the Helmholtz equation, *J. Comput. Phys.* **230** (2011), 4403–4418.
- [100] M. Bollhöfer, M. J. Grote, and O. Schenk, Algebraic multilevel preconditioner for the Helmholtz equation in heterogeneous media, *SIAM J. Sci. Comput.* **31** (2009), 3781–3805.
- [101] X.-C. Cai and O. B. Widlund, Domain decomposition algorithms for indefinite elliptic problems, *SIAM J. Sci. Stat. Comput.* **13** (1992), 243–258.
- [102] M. J. Gander and H. Zhang, *Domain decomposition methods for the Helmholtz equation: A numerical investigation*, in *Domain Decomposition Methods in Science and Engineering*, Vol **XX**, R. Bank et al., Eds., Springer-Verlag, Berlin/Heidelberg, Germany, 2013, 215–222.
- [103] C. Farhat, A. Macedo, and M. Lesoinne, A two-level domain decomposition method for the iterative solution of high frequency exterior Helmholtz problems, *Numer. Math.* **85** (2000), 283–308.

- [104] J.-D. Benamou and B. Desprès, A domain decomposition method for the Helmholtz equation and related optimal control problems, *J. Comput. Phys.* **136** (1997), 68–82.
- [105] M. J. Gander, F. Magoulès, and F. Nataf, Optimized Schwarz methods without overlap for the Helmholtz equation, *SIAM J. Sci. Comput.* **24** (2002), 38–60.
- [106] S. Kim, Parallel multidomain iterative algorithms for the Helmholtz wave equation, *Appl. Numer. Math.* **17** (1995), 411–429.
- [107] A. Bayliss, C. I. Goldstein, and E. Turkel, An iterative method for the Helmholtz equation, *J. Comput. Phys.* **49** (1983), 443–457.
- [108] A. L. Laird and M. B. Giles, *Preconditioned iterative solution of the 2D Helmholtz equation. Technical report NA-02-12*, University of Oxford, Oxford, UK, 2002.
- [109] M. Magolu Monga Made, R. Beauwens, and G. Warzée, Preconditioning of discrete Helmholtz operators perturbed by a diagonal complex matrix, *Commun. Numer. Methods Eng.* **16** (2000), 801–817.
- [110] Y. A. Erlangga, C. W. Oosterlee, and C. Vuik, A novel multigrid based preconditioner for heterogeneous Helmholtz problems, *SIAM J. Sci. Comput.* **27** (2006), 1471–1492.
- [111] A. H. Sheikh, D. Lahaye, and C. Vuik, On the convergence of shifted Laplace preconditioner combined with multilevel deflation, *Numer. Linear Algebra Appl.* **20** (2013), 645–662.
- [112] P.-H. Cocquet and M. J. Gander, How large a shift is needed in the shifted Helmholtz preconditioner for its effective inversion by multigrid? *SIAM J. Sci. Comput.* **39** (2017), A438–A478.
- [113] M. J. Gander, I. G. Graham, and E. A. Spence, Applying GMRES to the Helmholtz equation with shifted Laplacian preconditioning: What is the largest shift for which wavenumber-independent convergence is guaranteed? *Numer. Math.* **131** (2015), 567–614.
- [114] O. Axelsson, J. Karátson, and F. Magoulès, Superlinear convergence using block preconditioners for the real system formulation of complex Helmholtz equations, *J. Comput. Appl. Math.* **340** (2018), 424–431.
- [115] M. J. Gander and F. Nataf, An incomplete LU preconditioner for problems in acoustics, *J. Comput. Acoust.* **13** (2005), 455–476.
- [116] B. Engquist and L. Ying, Sweeping preconditioner for the Helmholtz equation: Moving perfectly matched layers, *Multiscale Model. Simul.* **9** (2011), 686–710.
- [117] M. J. Gander and H. Zhang, A class of iterative solvers for the Helmholtz equation: Factorizations, sweeping preconditioners, source transfer, single layer potentials, polarized traces, and optimized Schwarz methods, *SIAM Rev.* **61** (2019), 3–76.
- [118] X. Liu et al., Solving the three-dimensional high-frequency Helmholtz equation using contour integration and polynomial preconditioning, *SIAM J. Matrix Anal. Appl.* **41** (2020), 58–82.
- [119] Y. A. Erlangga, Advances in iterative methods and preconditioners for the Helmholtz equation, *Arch. Comput. Methods Eng.* **15** (2008), 37–66.
- [120] O. G. Ernst and M. J. Gander, *Why it is difficult to solve Helmholtz problems with classical iterative methods*, in *Numerical Analysis of Multiscale Problems*, I. G. Graham et al., Eds., Springer-Verlag, Berlin/Heidelberg, Germany, 2012, 325–363.
- [121] L. N. Trefethen and M. Embree, *Spectra and pseudospectra: The behavior of nonnormal matrices and operators*, Princeton University Press, Princeton, NJ, 2005.
- [122] O. Axelsson and J. Karátson, Symmetric part preconditioning for the conjugate gradient method in Hilbert space, *Numer. Funct. Anal. Optim.* **24** (2003), 455–474.
- [123] H. C. Elman and M. H. Schultz, Preconditioning by fast direct methods for nonself-adjoint nonseparable elliptic equations, *SIAM J. Numer. Anal.* **23** (1986), 44–57.
- [124] Z.-Z. Bai, G. H. Golub, and M. K. Ng, Hermitian and skew-Hermitian splitting methods for non-Hermitian positive definite linear systems, *SIAM J. Matrix Anal. Appl.* **24** (2003), 603–626.
- [125] D. Bertaccini et al., Preconditioned HSS methods for the solution of non-Hermitian positive definite linear systems and applications to the discrete convection-diffusion equation, *Numer. Math.* **99** (2005), 441–484.
- [126] R. C. Y. Chin and T. A. Manteuffel, An analysis of block successive overrelaxation for a class of matrices with complex spectra, *SIAM J. Numer. Anal.* **25** (1988), 564–585.
- [127] H. C. Elman and M. P. Chernesky, Ordering effects on relaxation methods applied to the discrete one-dimensional convection-diffusion equation, *SIAM J. Numer. Anal.* **30** (1993), 1268–1290.
- [128] H. C. Elman, D. J. Silvester, and A. J. Wathen, *Finite elements and fast iterative solvers with applications in incompressible fluid dynamics*, 2nd ed., Oxford University Press, Oxford, UK, 2014.
- [129] H. C. Elman and G. H. Golub, Iterative methods for cyclically reduced non-self-adjoint linear systems. II, *Math. Comput.* **56** (1991), 215–242.
- [130] H. Han et al., Analysis of flow directed iterations, *J. Comput. Math.* **10** (1992), 57–76.
- [131] P. A. Farrell, Flow conforming iterative methods for convection dominated flows, *IMACS Ann. Comput. Appl. Math.* **1** (1989), 681–686.
- [132] O. Axelsson and J. Karátson, Mesh independent superlinear PCG rates via compact-equivalent operators, *SIAM J. Numer. Anal.* **45** (2007), 1495–1516.
- [133] R. C. Y. Chin, T. A. Manteuffel, and J. de Pillis, ADI as a preconditioning for solving the convection-diffusion equation, *SIAM J. Sci. Stat. Comput.* **5** (1984), 281–299.
- [134] E. L. Wachspress, Extended application of alternating direction implicit iteration model problem theory, *J. Soc. Ind. Appl. Math.* **11** (1963), 994–1016.
- [135] D. Palitta and V. Simoncini, Matrix-equation-based strategies for convection–diffusion equations, *BIT Numer. Math.* **56** (2016), 751–776.
- [136] W. Hackbusch, A sparse matrix arithmetic based on \mathcal{H} -matrices. Part I: Introduction to \mathcal{H} -matrices, *Computing* **62** (1999), 89–108.

- [137] J. Bey and G. Wittum, Downwind numbering: Robust multigrid for convection-diffusion problems, *Appl. Numer. Math.* **23** (1997), 177–192.
- [138] W. Hackbusch and T. Probst, Downwind Gauß–Seidel smoothing for convection dominated problems, *Numer. Linear Algebra Appl.* **4** (1997), 85–102.
- [139] W. A. Mulder, A new multigrid approach to convection problems, *J. Comput. Phys.* **83** (1989), 303–323.
- [140] T. Washio and C. W. Oosterlee, Flexible multiple semicoarsening for three-dimensional singularly perturbed problems, *SIAM J. Sci. Comput.* **19** (1998), 1646–1666.
- [141] A. Brandt and I. Yavneh, On multigrid solution of high-Reynolds incompressible entering flows, *J. Comput. Phys.* **101** (1992), 151–164.
- [142] A. Ramage, A multigrid preconditioner for stabilised discretisations of advection-diffusion problems, *J. Comput. Appl. Math.* **110** (1999), 187–203.
- [143] M. Benzi and M. A. Olshanskii, An augmented Lagrangian-based approach to the Oseen problem, *SIAM J. Sci. Comput.* **28** (2006), 2095–2113.
- [144] P. E. Farrell, L. Mitchell, and F. Wechsung, An augmented Lagrangian preconditioner for the 3D stationary incompressible Navier–Stokes equations at high Reynolds number, *SIAM J. Sci. Comput.* **41** (2019), A3073–A3096.
- [145] J. Boyle, M. Mihajlović, and J. Scott, HSL_MI20: An efficient AMG preconditioner for finite element problems in 3D, *Int. J. Numer. Methods Eng.* **82** (2010), 64–98.
- [146] Y. Notay, Aggregation-based algebraic multigrid for convection-diffusion equations, *SIAM J. Sci. Comput.* **34** (2012), A2288–A2316.
- [147] C.-T. Wu and H. C. Elman, Analysis and comparison of geometric and algebraic multigrid for convection-diffusion equations, *SIAM J. Sci. Comput.* **28** (2006), 2208–2228.
- [148] V. Dolean, F. Nataf, and G. Rapin, Deriving a new domain decomposition method for the Stokes equations using the Smith factorization, *Math. Comput.* **78** (2009), 789–814.
- [149] J. Pestana et al., Efficient block preconditioning for a C^1 finite element discretization of the Dirichlet biharmonic problem, *SIAM J. Sci. Comput.* **38** (2016), A325–A345.
- [150] P. Farrell and J. Pestana, Block preconditioners for linear systems arising from multiscale collocation with compactly supported RBFs, *Numer. Linear Algebra Appl.* **22** (2015), 731–747.
- [151] M. Benzi, G. H. Golub, and J. Liesen, Numerical solution of saddle point problems, *Acta Numer.* **14** (2005), 1–137.
- [152] M. Benzi et al., A relaxed dimensional factorization preconditioner for the incompressible Navier–Stokes equations, *J. Comput. Phys.* **230** (2011), 6185–6202.
- [153] M. Benzi et al., Parameter estimates for the relaxed dimensional factorization preconditioner and application to hemodynamics, *Comput. Meth. Appl. Mech. Eng.* **300** (2016), 129–145.
- [154] M. F. Murphy, G. H. Golub, and A. J. Wathen, A note on preconditioning for indefinite linear systems, *SIAM J. Sci. Comput.* **21** (2000), 1969–1972.
- [155] N. Bootland et al., Preconditioners for two-phase incompressible Navier–Stokes flow, *SIAM J. Sci. Comput.* **41** (2019), B843–B869.
- [156] H. C. Elman, Preconditioning for the steady-state Navier–Stokes equations with low viscosity, *SIAM J. Sci. Comput.* **20** (1999), 1299–1316.
- [157] H. Elman et al., Block preconditioners based on approximate commutators, *SIAM J. Sci. Comput.* **27** (2006), 1651–1668.
- [158] T. Roy et al., A block preconditioner for non-isothermal flow in porous media, *J. Comput. Phys.* **395** (2019), 636–652.
- [159] J. A. White and R. I. Borja, Block-preconditioned Newton–Krylov solvers for fully coupled flow and geomechanics, *Comput. Geosci.* **15** (2011), 647–659.
- [160] Yu. A. Kuznetsov, Efficient iterative solvers for elliptic finite element problems on nonmatching grids, *Russ. J. Numer. Anal. Math. Model.* **10** (1995), 187–211.
- [161] G. H. Golub and R. S. Varga, Chebyshev semi-iterative methods, successive overrelaxation iterative methods, and second order Richardson iterative methods: Part I, *Numer. Math.* **3** (1961), 147–156.
- [162] G. H. Golub and R. S. Varga, Chebyshev semi-iterative methods, successive overrelaxation iterative methods, and second order Richardson iterative methods: Part II, *Numer. Math.* **3** (1961), 157–168.
- [163] A. Wathen and T. Rees, Chebyshev semi-iteration in preconditioning for problems including the mass matrix, *Electron. Trans. Numer. Anal.* **34** (2009), 125–135.
- [164] F. P. Ali Beik and M. Benzi, Iterative methods for double saddle point systems, *SIAM J. Matrix Anal. Appl.* **39** (2018), 902–921.
- [165] G. N. Gatica and N. Heuer, A dual-dual formulation for the coupling of mixed-FEM and BEM in hyperelasticity, *SIAM J. Numer. Anal.* **38** (2000), 380–400.
- [166] K.-A. Mardal, B. F. Nielsen, and M. Nordaas, Robust preconditioners for PDE-constrained optimization with limited observations, *BIT Numer. Math.* **57** (2017), 405–431.
- [167] J. Sogn and W. Zulehner, Schur complement preconditioners for multiple saddle point problems of block tridiagonal form with application to optimization problems, *IMA J. Numer. Anal.* **39** (2019), 1328–1359.
- [168] C. Powell and D. Silvester, *Black-box preconditioning for mixed formulation of self-adjoint elliptic PDEs*, in *Challenges in Scientific Computing – CISC 2002*, E. Bänsch, Ed., Springer-Verlag, Berlin/Heidelberg, Germany, 2003, 268–285.
- [169] J. Sundnes et al., Multigrid block preconditioning for a coupled system of partial differential equations modeling the electrical activity in the heart, *Comput. Methods Biomech. Biomed. Engin.* **5** (2002), 397–409.
- [170] D. Silvester and A. Wathen, Fast iterative solution of stabilised Stokes systems Part II: Using general block preconditioners, *SIAM J. Numer. Anal.* **31** (1994), 1352–1367.

- [171] P. S. Vassilevski and U. Villa, A block-diagonal algebraic multigrid preconditioner for the Brinkman problem, *SIAM J. Sci. Comput.* **35** (2013), S3–S17.
- [172] B. Fischer et al., Minimum residual methods for augmented systems, *BIT Numer. Math.* **38** (1998), 527–543.
- [173] J. W. Pearson, J. Pestana, and D. J. Silvester, Refined saddle-point preconditioners for discretized Stokes problems, *Numer. Math.* **138** (2018), 331–363.
- [174] R. Herzog, Dimensionally consistent preconditioning for saddle-point problems, arXiv e-prints. 2020: arXiv:2003.09478.
- [175] P. P. Grinevich and M. A. Olshanskii, An iterative method for the Stokes-type problem with variable viscosity, *SIAM J. Sci. Comput.* **31** (2009), 3959–3978.
- [176] S. Holmgren and K. Otto, Semicirculant preconditioners for first-order partial differential equations, *SIAM J. Sci. Comput.* **15** (1994), 385–407.
- [177] K. Otto, Analysis of preconditioners for hyperbolic partial differential equations, *SIAM J. Numer. Anal.* **33** (1996), 2131–2165.
- [178] M. M. Meerschaert and C. Tadjeran, Finite difference approximations for fractional advection–dispersion flow equations, *J. Comput. Appl. Math.* **172** (2004), 65–77.
- [179] M. M. Meerschaert and C. Tadjeran, Finite difference approximations for two-sided space-fractional partial differential equations, *Appl. Numer. Math.* **56** (2006), 80–90.
- [180] G. H. Golub and C. F. Van Loan, *Matrix computations*, 4th ed., The John Hopkins University Press, Maryland, 2013.
- [181] U. Grenander and G. Szegő, *Toeplitz forms and their applications*, American Mathematical Society, New York, 1958.
- [182] E. E. Tyrtyshnikov, A unifying approach to some old and new theorems on distribution and clustering, *Linear Algebra Appl.* **232** (1996), 1–43.
- [183] N. L. Zamarashkin and E. E. Tyrtyshnikov, Distribution of eigenvalues and singular values of Toeplitz matrices under weakened conditions on the generating function, *Sb. Math.* **188** (1997), 1191–1201.
- [184] P. Tilli, A note on the spectral distribution of Toeplitz matrices, *Linear Multilinear Algebra* **45** (1998), 147–159.
- [185] G. Strang, A proposal for Toeplitz matrix calculations, *Stud. Appl. Math.* **74** (1986), 171–176.
- [186] J. A. Olkin, *Linear and nonlinear deconvolution problems*, PhD Thesis, Rice University, 1986.
- [187] T. F. Chan, An optimal circulant preconditioner for Toeplitz systems, *SIAM J. Sci. Stat. Comput.* **9** (1988), 766–771.
- [188] E. E. Tyrtyshnikov, Optimal and superoptimal circulant preconditioners, *SIAM J. Matrix Anal. Appl.* **13** (1992), 459–473.
- [189] D. Bini and P. Favati, On a matrix algebra related to the discrete Hartley transform, *SIAM J. Matrix Anal. Appl.* **14** (1993), 500–507.
- [190] T. Huckle, Fast transforms for tridiagonal linear equations, *BIT Numer. Math.* **34** (1994), 99–112.
- [191] R. H. Chan, Toeplitz preconditioners for Toeplitz systems with nonnegative generating functions, *IMA J. Numer. Anal.* **11** (1991), 333–345.
- [192] R. H. Chan and P. T. P. Tang, Fast band-Toeplitz preconditioners for Hermitian Toeplitz systems, *SIAM J. Sci. Comput.* **15** (1994), 164–171.
- [193] G. Fiorentino and S. Serra, Multigrid methods for Toeplitz matrices, *Calcolo* **28** (1991), 283–305.
- [194] R. H. Chan and M. K. Ng, Conjugate gradient methods for Toeplitz systems, *SIAM Rev.* **38** (1996), 427–482.
- [195] R. H.-F. Chan and X.-Q. Jin, *An introduction to iterative Toeplitz solvers*, Society for Industrial and Applied Mathematics, Philadelphia, PA, 2007.
- [196] M. K. Ng, *Iterative methods for Toeplitz systems*, Oxford University Press, Oxford, UK, 2004.
- [197] S. Serra Capizzano and E. Tyrtyshnikov, How to prove that a preconditioner cannot be superlinear, *Math. Comput.* **72** (2003), 1305–1316.
- [198] S. Serra Capizzano and E. Tyrtyshnikov, Any circulant-like preconditioner for multilevel matrices is not superlinear, *SIAM J. Matrix Anal. Appl.* **21** (1999), 431–439.
- [199] A. Aricò and M. Donatelli, A V-cycle multigrid for multilevel matrix algebras: Proof of optimality, *Numer. Math.* **105** (2007), 511–547.
- [200] S. Serra, Spectral and computational analysis of block Toeplitz matrices having nonnegative definite matrix-valued generating functions, *BIT Numer. Math.* **39** (1999), 152–175.
- [201] C. Garoni and S. Serra-Capizzano, *Generalized locally Toeplitz sequences: Theory and applications*, Volume I, Springer, New York, NY, 2017.
- [202] C. Garoni and S. Serra-Capizzano, *Generalized locally Toeplitz sequences: Theory and applications*, Volume II, Springer, New York, NY, 2018.
- [203] M. J. Gander, 50 years of time parallel time integration, in *Multiple Shooting and Time Domain Decomposition Methods*, T. Carraro et al., Eds., Springer, New York, NY, 2015, 69–113.
- [204] D. Palitta, Matrix equation techniques for certain evolutionary partial differential equations, arXiv e-prints. 2019: arXiv:1908.11851.
- [205] E. McDonald, J. Pestana, and A. Wathen, Preconditioning and iterative solution of all-at-once systems for evolutionary partial differential equations, *SIAM J. Sci. Comput.* **40** (2018), A1012–A1033.
- [206] A. Goddard and A. Wathen, A note on parallel preconditioning for all-at-once evolutionary PDEs, *Electron. Trans. Numer. Anal.* **51** (2019), 135–150.
- [207] S. Güttel and J. W. Pearson, A rational deferred correction approach to parabolic optimal control problems, *IMA J. Numer. Anal.* **38** (2018), 1861–1892.
- [208] M. Benzi, E. Haber, and L. Taralli, A preconditioning technique for a class of PDE-constrained optimization problems, *Adv. Comput. Math.* **35** (2011), 149–173.
- [209] M. Hinze, M. Köster, and S. Turek, *A space-time multigrid method for optimal flow control*, *Internat. Ser. Numer. Math.*, Vol **160**, Springer, Basel, AG, 2012, 147–170.

- [210] M. Stoll and A. Wathen, *All-at-once solution of time-dependent PDE-constrained optimization problems. Technical report NA-10-13*, University of Oxford, Oxford, UK, 2010.
- [211] I. G. Graham and M. J. Hagger, Unstructured additive Schwarz–conjugate gradient method for elliptic problems with highly discontinuous coefficients, *SIAM J. Sci. Comput.* **20** (1999), 2041–2066.
- [212] P. Bastian, M. Blatt, and R. Scheichl, Algebraic multigrid for discontinuous Galerkin discretizations of heterogeneous elliptic problems, *Numer. Linear Algebra Appl.* **19** (2012), 367–388.
- [213] J. Kraus et al., Preconditioning heterogeneous $\mathbf{H}(\text{div})$ problems by additive Schur complement approximation and applications, *SIAM J. Sci. Comput.* **38** (2016), A875–A898.
- [214] A. Klawonn, O. B. Widlund, and M. Dryja, Dual-primal FETI methods for three-dimensional elliptic problems with heterogeneous coefficients, *SIAM J. Numer. Anal.* **40** (2002), 159–179.
- [215] C. E. Powell and D. Silvester, Optimal preconditioning for Raviart–Thomas mixed formulation of second-order elliptic problems, *SIAM J. Matrix Anal. Appl.* **25** (2004), 718–738.
- [216] M. W. Benson, Iterative solution of large scale linear systems, Master’s Thesis, Lakehead University Thunder Bay, Ontario, 1973.
- [217] G. Alléon, M. Benzi, and L. Giraud, Sparse approximate inverse preconditioning for dense linear systems arising in computational electromagnetics, *Numer. Alg.* **16** (1997), 1–15.
- [218] B. Carpentieri, I. S. Duff, and L. Giraud, Sparse pattern selection strategies for robust Frobenius-norm minimization preconditioners in electromagnetism, *Numer. Linear Algebra Appl.* **7** (2000), 667–685.
- [219] A. Franceschini et al., Recent advancements in preconditioning techniques for large size linear systems suited for High Performance Computing. *Dolomites Res. Notes Approx.* **32** (2018), 11–22.
- [220] L. Giraud, J. Langou, and G. Sylvand, On the parallel solution of large industrial wave propagation problems, *J. Comput. Acoust.* **14** (2006), 83–111.
- [221] L. Yu. Kolotilina and A. Yu. Yeremin, Factorized sparse approximate inverse preconditioning. I. Theory, *SIAM J. Matrix Anal. Appl.* **14** (1993), 45–58.
- [222] M. Margonari and M. Bonnet, Fast multipole method applied to elastostatic BEM–FEM coupling, *Comput. Struct.* **83** (2005), 700–717.
- [223] J. A. Sanz, M. Bonnet, and J. Dominguez, Fast multipole method applied to 3-D frequency domain elastodynamics, *Eng. Anal. Bound. Elem.* **32** (2008), 787–795.
- [224] M. J. Grote and T. Huckle, Parallel preconditioning with sparse approximate inverses, *SIAM J. Sci. Comput.* **18** (1997), 838–853.
- [225] M. Benzi and M. Tũma, A sparse approximate inverse preconditioner for nonsymmetric linear systems, *SIAM J. Sci. Comput.* **19** (1998), 968–994.
- [226] M. Benzi and M. Tũma, A comparative study of sparse approximate inverse preconditioners, *Appl. Numer. Math.* **30** (1999), 305–340.
- [227] T. Huckle, Approximate sparsity patterns for the inverse of a matrix and preconditioning, *Appl. Numer. Math.* **30** (1999), 291–303.
- [228] T. K. Huckle, Efficient computation of sparse approximate inverses, *Numer. Linear Algebra Appl.* **5** (1998), 57–71.
- [229] C. H. Ahn et al., Numerical study of approximate inverse preconditioner for two-dimensional engine inlet problems, *Electromagnetics* **19** (1999), 131–146.
- [230] B. Carpentieri et al., Sparse symmetric preconditioners for dense linear systems in electromagnetism, *Numer. Linear Algebra Appl.* **11** (2004), 753–771.
- [231] M. Wathen and C. Greif, A scalable approximate inverse block preconditioner for an incompressible magnetohydrodynamics model problem, *SIAM J. Sci. Comput.* **42** (2020), B57–B79.
- [232] D. W. Peaceman and H. H. Rachford Jr., The numerical solution of parabolic and elliptic differential equations, *J. Soc. Ind. Appl. Math.* **3** (1955), 28–41.
- [233] Z. Dostál, Conjugate gradient method with preconditioning by projector, *Int. J. Comput. Math.* **23** (1988), 315–323.
- [234] A. Gaul et al., A framework for deflated and augmented Krylov subspace methods, *SIAM J. Matrix Anal. Appl.* **34** (2013), 495–518.
- [235] M. Bebendorf, *Hierarchical matrices: A means to efficiently solve elliptic boundary value problems*, Springer-Verlag, Berlin/Heidelberg, Germany, 2008.
- [236] L. Grasedyck and W. Hackbusch, Construction and arithmetics of \mathcal{H} -matrices, *Computing* **70** (2003), 295–334.
- [237] W. Hackbusch, B. Khoromskij, and S. A. Sauter, On \mathcal{H}^2 -matrices, in *Lectures on Applied Mathematics*, H.-J. Bungartz, R. H. W. Hoppe, and C. Zenger, Eds., Springer-Verlag, Berlin/Heidelberg, Germany, 2000, 9–29.
- [238] S. Chandrasekaran et al., Some fast algorithms for sequentially semiseparable representations, *SIAM J. Matrix Anal. Appl.* **27** (2005), 341–364.
- [239] L. Greengard and V. Rokhlin, A fast algorithm for particle simulations, *J. Comput. Phys.* **73** (1987), 325–348.
- [240] K. L. Ho and L. Greengard, A fast direct solver for structured linear systems by recursive skeletonization, *SIAM J. Sci. Comput.* **34** (2012), A2507–A2532.
- [241] S. Le Borne and L. Grasedyck, \mathcal{H} -matrix preconditioners in convection-dominated problems, *SIAM J. Matrix Anal. Appl.* **27** (2006), 1172–1183.
- [242] H. Ibeid et al., Fast multipole preconditioners for sparse matrices arising from elliptic equations, *Comput. Vis. Sci.* **18** (2018), 213–229.
- [243] R. Bridson and C. Greif, A multipreconditioned conjugate gradient algorithm, *SIAM J. Matrix Anal. Appl.* **27** (2006), 1056–1068.
- [244] C. Greif, T. Rees, and D. B. Szyld, GMRES with multiple preconditioners, *SeMA Journal* **74** (2017), 213–231.
- [245] T. Bakhos et al., Multipreconditioned GMRES for shifted systems, *SIAM J. Sci. Comput.* **39** (2017), S222–S247.
- [246] N. Bootland, A. Wathen, Multipreconditioning with application to two-phase incompressible Navier-Stokes flow, *arXiv e-prints*. 2020: arXiv:2005.07608.

- [247] J. Gondzio, Interior point methods 25 years later, *Eur. J. Oper. Res.* **218** (2012), 587–601.
- [248] L. Bergamaschi, J. Gondzio, and G. Zilli, Preconditioning indefinite systems in interior point methods for optimization, *Comput. Optim. Appl.* **28** (2004), 149–171.
- [249] P. E. Gill et al., Preconditioners for indefinite systems arising in optimization, *SIAM J. Matrix Anal. Appl.* **13** (1992), 292–311.
- [250] A. R. L. Oliveira and D. C. Sorensen, A new class of preconditioners for large-scale linear systems from interior point methods for linear programming, *Linear Algebra Appl.* **394** (2005), 1–24.
- [251] L. Schork and J. Gondzio, Implementation of an interior point method with basis preconditioning, *Math. Prog. Comp.* (2020). <https://doi.org/10.1007/s12532-020-00181-8>.
- [252] T. A. Davis, Algorithm 832: UMFPACK V4.3—an unsymmetric-pattern multifrontal method, *ACM Trans. Math. Softw.* **30** (2004), 196–199.
- [253] X. S. Li, An overview of SuperLU: Algorithms, implementation, and user interface, *ACM Trans. Math. Softw.* **31** (2005), 302–325.
- [254] M. Bollhöfer, A robust ILU with pivoting based on monitoring the growth of the inverse factors, *Linear Algebra Appl.* **338** (2001), 201–218.
- [255] E. Chow and Y. Saad, Experimental study of ILU preconditioners for indefinite matrices, *J. Comput. Appl. Math.* **86** (1997), 387–414.
- [256] M. T. Jones and P. E. Plassmann, An improved incomplete Cholesky factorization, *ACM Trans. Math. Softw.* **21** (1995), 5–17.
- [257] Y. Notay, A multilevel block incomplete factorization preconditioning, *Appl. Numer. Math.* **31** (1999), 209–225.
- [258] Y. Saad, ILUM: A multi-elimination ILU preconditioner for general sparse matrices, *SIAM J. Sci. Comput.* **17** (1996), 830–847.
- [259] S. C. Eisenstat, Efficient implementation of a class of preconditioned conjugate gradient methods, *SIAM J. Sci. Stat. Comput.* **2** (1981), 1–4.
- [260] M. Benzi and M. Tüma, A robust incomplete factorization preconditioner for positive definite matrices, *Numer. Linear Algebra Appl.* **10** (2003), 385–400.
- [261] R. E. Bank and C. Wagner, Multilevel ILU decomposition, *Numer. Math.* **82** (1999), 543–576.
- [262] P. S. Vassilevski, *Multilevel block factorization preconditioners: Matrix-based analysis and algorithms for solving finite element equations*, Springer-Verlag, New York, NY, 2008.
- [263] J.-S. Chai and K.-C. Toh, Preconditioning and iterative solution of symmetric indefinite linear systems arising from interior point methods for linear programming, *Comput. Optim. Appl.* **36** (2007), 221–247.
- [264] J. C. Haws and C. D. Meyer, *Preconditioning KKT systems. Technical report M&CT-Tech-01-021*, The Boeing Company, Seattle, WA, 2003.
- [265] O. Schenk, A. Wächter, and M. Weiser, Inertia-revealing preconditioning for large-scale nonconvex constrained optimization, *SIAM J. Sci. Comput.* **31** (2008), 939–960.
- [266] N. I. M. Gould, M. E. Hribar, and J. Nocedal, On the solution of equality constrained quadratic programming problems arising in optimization, *SIAM J. Sci. Comput.* **23** (2001), 1376–1395.
- [267] C. Keller, N. I. M. Gould, and A. J. Wathen, Constraint preconditioning for indefinite linear systems, *SIAM J. Matrix Anal. Appl.* **21** (2000), 1300–1317.
- [268] L. Lukšan and J. Vlček, Indefinitely preconditioned inexact Newton method for large sparse equality constrained non-linear programming problems, *Numer. Linear Algebra Appl.* **5** (1998), 219–247.
- [269] N. Gould, D. Orban, and T. Rees, Projected Krylov methods for saddle-point systems, *SIAM J. Matrix Anal. Appl.* **35** (2014), 1329–1343.
- [270] H. S. Dollar et al., Implicit-factorization preconditioning and iterative solvers for regularized saddle-point systems, *SIAM J. Matrix Anal. Appl.* **28** (2006), 170–189.
- [271] H. S. Dollar et al., Using constraint preconditioners with regularized saddle-point problems, *Comput. Optim. Appl.* **36** (2007), 249–270.
- [272] T. Rees and J. Scott, A comparative study of null-space factorizations for sparse symmetric saddle point systems, *Numer. Linear Algebra Appl.* **25** (2018), e2103.
- [273] C. Durazzi and V. Ruggiero, Indefinitely preconditioned conjugate gradient method for large sparse equality and inequality constrained quadratic problems, *Numer. Linear Algebra Appl.* **10** (2003), 673–688.
- [274] L. Bergamaschi et al., Inexact constraint preconditioners for linear systems arising in interior point methods, *Comput. Optim. Appl.* **36** (2007), 137–147.
- [275] N. Dyn and W. E. Ferguson Jr., The numerical solution of equality-constrained quadratic programming problems, *Math. Comput.* **41** (1983), 165–170.
- [276] A. Forsgren, P. E. Gill, and J. D. Griffin, Iterative solution of augmented systems arising in interior point methods, *SIAM J. Optim.* **18** (2007), 666–690.
- [277] L. Lukšan, C. Matonoha, and J. Vlček, Interior-point method for non-linear non-convex optimization, *Numer. Linear Algebra Appl.* **11** (2004), 431–453.
- [278] M. Rozložník and V. Simoncini, Krylov subspace methods for saddle point problems with indefinite preconditioning, *SIAM J. Matrix Anal. Appl.* **24** (2002), 368–391.
- [279] E. de Sturler and J. Liesen, Block-diagonal and constraint preconditioners for nonsymmetric indefinite linear systems. Part I: Theory, *SIAM J. Sci. Comput.* **26** (2005), 1598–1619.
- [280] J. Scott and M. Tüma, Improving the stability and robustness of incomplete symmetric indefinite factorization preconditioners, *Numer. Linear Algebra Appl.* **24** (2017), e2099.
- [281] T. Rees, H. S. Dollar, and A. J. Wathen, Optimal solvers for PDE-constrained optimization, *SIAM J. Sci. Comput.* **32** (2010), 271–298.
- [282] M. Kočvara, D. Loghin, and J. Turner, Constraint interface preconditioning for topology optimization problems, *SIAM J. Sci. Comput.* **38** (2016), A128–A145.

- [283] C. Greif and D. Schötzau, Preconditioners for saddle point linear systems with highly singular (1,1) blocks, *Electron. Trans. Numer. Anal.* **22** (2006), 114–121.
- [284] C. Greif and D. Schötzau, Preconditioners for the discretized time-harmonic Maxwell equations in mixed form, *Numer. Linear Algebra Appl.* **14** (2007), 281–297.
- [285] G. H. Golub and C. Greif, On solving block-structured indefinite linear systems, *SIAM J. Sci. Comput.* **24** (2003), 2076–2092.
- [286] R. Fletcher, *Practical methods of optimization*, 2nd ed., John Wiley & Sons, Chichester, 1987.
- [287] M. Fortin and R. Glowinski, *Augmented Lagrangian methods: Applications to the solution of numerical boundary-value problems*, Vol **15**, North-Holland, Amsterdam, Netherlands, 1983.
- [288] T. Rees and C. Greif, A preconditioner for linear systems arising from interior point optimization methods, *SIAM J. Sci. Comput.* **29** (2007), 1992–2007.
- [289] B. Morini, V. Simoncini, and M. Tani, Spectral estimates for unreduced symmetric KKT systems arising from Interior Point methods, *Numer. Linear Algebra Appl.* **23** (2016), 776–800.
- [290] Z.-H. Cao, Augmentation block preconditioners for saddle point-type matrices with singular (1,1) blocks, *Numer. Linear Algebra Appl.* **15** (2008), 515–533.
- [291] S.-Q. Shen, T.-Z. Huang, and J.-S. Zhang, Augmentation block triangular preconditioners for regularized saddle point problems, *SIAM J. Matrix Anal. Appl.* **33** (2012), 721–741.
- [292] L. Bergamaschi et al., Quasi-Newton preconditioners for the inexact Newton method, *Electron. Trans. Numer. Anal.* **23** (2006), 76–87.
- [293] J. M. Martínez, A theory of secant preconditioners, *Math. Comput.* **60** (1993), 681–698.
- [294] J. M. Martínez, An extension of the theory of secant preconditioners, *J. Comput. Appl. Math.* **60** (1995), 115–125.
- [295] J. L. Morales and J. Nocedal, Automatic preconditioning by limited memory quasi-Newton updating, *SIAM J. Optim.* **10** (2000), 1079–1096.
- [296] S. G. Nash, Newton-type minimization via the Lanczos method, *SIAM J. Numer. Anal.* **21** (1984), 770–788.
- [297] D. P. O’Leary and A. Yeregin, The linear algebra of block quasi-Newton updates, *Linear Algebra Appl.* **212–213** (1994), 153–168.
- [298] L. Bergamaschi, R. Bru, and A. Martínez, Low-rank update of preconditioners for the inexact Newton method with SPD Jacobian, *Math. Comput. Model.* **54** (2011), 1863–1873.
- [299] J. Duintjer Tebbens and M. Tũma, Efficient preconditioning of sequences of nonsymmetric linear systems, *SIAM J. Sci. Comput.* **29** (2007), 1918–1941.
- [300] J. Duintjer Tebbens and M. Tũma, Preconditioner updates for solving sequences of linear systems in matrix-free environment, *Numer. Linear Algebra Appl.* **17** (2010), 997–1019.
- [301] S. Bellavia et al., Updating constraint preconditioners for KKT systems in quadratic programming via low-rank corrections, *SIAM J. Optim.* **25** (2015), 1787–1808.
- [302] S. Bellavia et al., On the update of constraint preconditioners for regularized KKT systems, *Comput. Optim. Appl.* **65** (2016), 339–360.
- [303] L. Bergamaschi et al., BFGS-like updates of constraint preconditioners for sequences of KKT linear systems in quadratic programming, *Numer. Linear Algebra Appl.* **25** (2018), e2144.
- [304] S. Gratton, A. Sartenaer, and J. Tshimanga, On a class of limited memory preconditioners for large scale linear systems with multiple right-hand sides, *SIAM J. Optim.* **21** (2011), 912–935.
- [305] M. Fisher et al., Low rank updates in preconditioning the saddle point systems arising from data assimilation problems, *Optim. Method. Softw.* **33** (2018), 45–69.
- [306] S. Bellavia et al., Efficient preconditioner updates for shifted linear systems, *SIAM J. Sci. Comput.* **33** (2011), 1785–1809.
- [307] M. Benzi and D. Bertaccini, Approximate inverse preconditioning for shifted linear systems, *BIT Numer. Math.* **43** (2003), 231–244.
- [308] X.-M. Gu et al., Restarted Hessenberg method for solving shifted nonsymmetric linear systems, *J. Comput. Appl. Math.* **331** (2018), 166–177.
- [309] L. Bergamaschi, A survey of low-rank updates of preconditioners for sequences of symmetric linear systems, *Algorithms* **13** (2020), 100.
- [310] R. Bru et al., Preconditioned iterative methods for solving linear least squares problems, *SIAM J. Sci. Comput.* **36** (2014), A2002–A2022.
- [311] Gould N., Scott J., The state-of-the-art of preconditioners for sparse linear least-squares problems. *ACM Trans. Math. Softw.* 2017;**43**:Art. 36.
- [312] Scott J., Tũma M., HSL_MI28: An efficient and robust limited-memory incomplete Cholesky factorization code. *ACM Trans. Math. Softw.* 2014;**40**:Art. 24.
- [313] J. Scott and M. Tũma, On positive semidefinite modification schemes for incomplete Cholesky factorization, *SIAM J. Sci. Comput.* **36** (2014), A609–A633.
- [314] J. Scott, On using Cholesky-based factorizations and regularization for solving rank-deficient sparse linear least-squares problems, *SIAM J. Sci. Comput.* **39** (2017), C319–C339.
- [315] Z.-Z. Bai, I. S. Duff, and A. J. Wathen, A class of incomplete orthogonal factorization methods. I: Methods and theories, *BIT Numer. Math.* **41** (2001), 53–70.
- [316] A. T. Papadopoulos, I. S. Duff, and A. J. Wathen, A class of incomplete orthogonal factorization methods. II: Implementation and results, *BIT Numer. Math.* **45** (2005), 159–179.
- [317] A. Jennings and M. A. Ajiz, Incomplete methods for solving $A^T Ax = b$, *SIAM J. Sci. Stat. Comput.* **5** (1984), 978–987.
- [318] X. Wang, K. A. Gallivan, and R. Bramley, CIMGS: An incomplete orthogonal factorization preconditioner, *SIAM J. Sci. Comput.* **18** (1997), 516–536.

- [319] N. Li and Y. Saad, MIQR: A multilevel incomplete QR preconditioner for large sparse least-squares problems, *SIAM J. Matrix Anal. Appl.* **28** (2006), 524–550.
- [320] M. Benzi and M. Tũma, A robust preconditioner with low memory requirements for large sparse least squares problems, *SIAM J. Sci. Comput.* **25** (2003), 499–512.
- [321] J. Scott and M. Tũma, On signed incomplete Cholesky factorization preconditioners for saddle-point systems, *SIAM J. Sci. Comput.* **36** (2014), A2984–A3010.
- [322] Greif C., He S., Liu P., SYM-ILDL: Incomplete LDL^T factorization of symmetric indefinite and skew-symmetric matrices. *ACM Trans. Math. Softw.* 2017;**44**:Art. 1.
- [323] X. Cui and K. Hayami, Generalized approximate inverse preconditioners for least squares problems, *Jpn. J. Ind. Appl. Math.* **26** (2009), 1–14.
- [324] X. Cui, K. Hayami, and J.-F. Yin, Greville’s method for preconditioning least squares problems, *Adv. Comput. Math.* **35** (2011), 243–269.
- [325] A. Björck, *Numerical methods for least squares problems*, Society for Industrial and Applied Mathematics, Philadelphia, PA, 1996.
- [326] A. Björck and J. Y. Yuan, Preconditioners for least squares problems by LU factorization, *Electron. Trans. Numer. Anal.* **8** (1999), 26–35.
- [327] M. Arioli and I. S. Duff, Preconditioning linear least-squares problems by identifying a basis matrix, *SIAM J. Sci. Comput.* **37** (2015), S544–S561.
- [328] M. Benzi and M. K. Ng, Preconditioned iterative methods for weighted Toeplitz least squares problems, *SIAM J. Matrix Anal. Appl.* **27** (2006), 1106–1124.
- [329] S. Börm and S. Le Borne, \mathcal{H} -LU factorization in preconditioners for augmented Lagrangian and grad-div stabilized saddle point systems, *Int. J. Numer. Meth. Fl.* **68** (2012), 83–98.
- [330] Y. A. Erlangga, C. Vuik, and C. W. Oosterlee, Comparison of multigrid and incomplete LU shifted-Laplace preconditioners for the inhomogeneous Helmholtz equation, *Appl. Numer. Math.* **56** (2006), 648–666.
- [331] Z.-Z. Bai, M. K. Ng, and Z.-Q. Wang, Constraint preconditioners for symmetric indefinite matrices, *SIAM J. Matrix Anal. Appl.* **31** (2009), 410–433.
- [332] P. Chidyagwai, S. Ladenheim, and D. B. Szyld, Constraint preconditioning for the coupled Stokes–Darcy system, *SIAM J. Sci. Comput.* **38** (2016), A668–A690.
- [333] I. Perugia and V. Simoncini, Block-diagonal and indefinite symmetric preconditioners for mixed finite element formulations, *Numer. Linear Algebra Appl.* **7** (2000), 585–616.
- [334] M. Benzi, M. A. Olshanskii, and Z. Wang, Modified augmented Lagrangian preconditioners for the incompressible Navier–Stokes equations, *Int. J. Numer. Meth. Fl.* **66** (2011), 486–508.
- [335] J. W. Pearson and A. J. Wathen, A new approximation of the Schur complement in preconditioners for PDE-constrained optimization problems, *Numer. Linear Algebra Appl.* **19** (2012), 816–829.
- [336] J. Schöberl and W. Zulehner, Symmetric indefinite preconditioners for saddle point problems with applications to PDE-constrained optimization problems, *SIAM J. Matrix Anal. Appl.* **29** (2007), 752–773.
- [337] J. Pearson, Fast iterative solvers for PDE-constrained optimization problems, PhD Thesis, University of Oxford, 2013.
- [338] T. Rees and M. Stoll, Block-triangular preconditioners for PDE-constrained optimization, *Numer. Linear Algebra Appl.* **17** (2010), 977–996.
- [339] W. Krendl, V. Simoncini, and W. Zulehner, Stability estimates and structural spectral properties of saddle point problems, *Numer. Math.* **124** (2013), 183–213.
- [340] J. W. Pearson and A. J. Wathen, Fast iterative solvers for convection–diffusion control problems, *Electron. Trans. Numer. Anal.* **40** (2013), 294–310.
- [341] J. W. Pearson, M. Stoll, and A. J. Wathen, Regularization-robust preconditioners for time-dependent PDE-constrained optimization problems, *SIAM J. Matrix Anal. Appl.* **33** (2012), 1126–1152.
- [342] Z.-Z. Bai et al., Preconditioned MHSS iteration methods for a class of block two-by-two linear systems with applications to distributed control problems, *IMA J. Numer. Anal.* **33** (2013), 343–369.
- [343] O. Axelsson, S. Farouq, and M. Neytcheva, Comparison of preconditioned Krylov subspace iteration methods for PDE-constrained optimization problems: Poisson and convection–diffusion control, *Numer. Alg.* **73** (2016), 631–663.
- [344] R. Herzog and E. Sachs, Preconditioned conjugate gradient method for optimal control problems with control and state constraints, *SIAM J. Matrix Anal. Appl.* **31** (2010), 2291–2317.
- [345] R. Herzog and S. Mach, Preconditioned solution of state gradient constrained elliptic optimal control problems, *SIAM J. Numer. Anal.* **54** (2016), 688–718.
- [346] J. W. Pearson and J. Gondzio, Fast interior point solution of quadratic programming problems arising from PDE-constrained optimization problems, *Numer. Math.* **137** (2017), 959–999.
- [347] J. W. Pearson, M. Stoll, and A. J. Wathen, Preconditioners for state constrained optimal control problems with Moreau–Yosida penalty function, *Numer. Linear Algebra Appl.* **21** (2012), 81–97.
- [348] M. Porcelli, V. Simoncini, and M. Tani, Preconditioning of active-set Newton methods for PDE-constrained optimal control problems, *SIAM J. Sci. Comput.* **37** (2015), S472–S502.
- [349] E. Haber and U. M. Ascher, Preconditioned all-at-once methods for large, sparse parameter estimation problems, *Inverse Probl.* **17** (2001), 1847–1864.
- [350] J. Berns-Müller, I. G. Graham, and A. Spence, Inexact inverse iteration for symmetric matrices, *Linear Algebra Appl.* **416** (2006), 389–413.

- [351] G. H. Golub and Q. Ye, Inexact preconditioned conjugate gradient method with inner-outer iteration, *SIAM J. Sci. Comput.* **21** (1999), 1305–1320.
- [352] M. A. Freitag and A. Spence, Convergence theory for inexact inverse iteration applied to the generalised nonsymmetric eigenproblem, *Electron. Trans. Numer. Anal.* **28** (2007), 40–64.
- [353] M. A. Freitag and A. Spence, Convergence of inexact inverse iteration with application to preconditioned iterative solves, *BIT Numer. Math.* **47** (2007), 27–44.
- [354] V. Simoncini and L. Eldén, Inexact Rayleigh quotient-type methods for eigenvalue computations, *BIT Numer. Math.* **42** (2002), 159–182.
- [355] M. A. Freitag and A. Spence, A tuned preconditioner for inexact inverse iteration applied to Hermitian eigenvalue problems, *IMA J. Numer. Anal.* **28** (2008), 522–551.
- [356] D. B. Szyld and F. Xue, Efficient preconditioned inner solves for inexact Rayleigh quotient iteration and their connections to the single-vector Jacobi–Davidson method, *SIAM J. Matrix Anal. Appl.* **32** (2011), 993–1018.
- [357] M. Robbé, M. Sadkane, and A. Spence, Inexact inverse subspace iteration with preconditioning applied to non-Hermitian eigenvalue problems, *SIAM J. Matrix Anal. Appl.* **31** (2009), 92–113.
- [358] M. A. Freitag and A. Spence, Shift-invert Arnoldi’s method with preconditioned iterative solves, *SIAM J. Matrix Anal. Appl.* **31** (2009), 942–969.
- [359] F. Xue and H. C. Elman, Fast inexact implicitly restarted Arnoldi method for generalized eigenvalue problems with spectral transformation, *SIAM J. Matrix Anal. Appl.* **33** (2012), 433–459.
- [360] G. H. Golub and Q. Ye, An inverse free preconditioned Krylov subspace method for symmetric generalized eigenvalue problems, *SIAM J. Sci. Comput.* **24** (2002), 312–334.
- [361] R. B. Morgan and D. S. Scott, Preconditioning the Lanczos algorithm for sparse symmetric eigenvalue problems, *SIAM J. Sci. Comput.* **14** (1993), 585–593.
- [362] L. Wu, F. Xue, and A. Stathopoulos, TRPL+K: Thick-restart preconditioned Lanczos+K method for large symmetric eigenvalue problems, *SIAM J. Sci. Comput.* **41** (2019), A1013–A1040.
- [363] O. Taussky, The role of symmetric matrices in the study of general matrices, *Linear Algebra Appl.* **5** (1972), 147–154.
- [364] S. Hon and A. Wathen, Circulant preconditioners for analytic functions of Toeplitz matrices, *Numer. Alg.* **79** (2018), 1211–1230.
- [365] J. Pestana and A. J. Wathen, A preconditioned MINRES method for nonsymmetric Toeplitz matrices, *SIAM J. Matrix Anal. Appl.* **36** (2015), 273–288.
- [366] J. Pestana, Preconditioners for symmetrized Toeplitz and multilevel Toeplitz matrices, *SIAM J. Matrix Anal. Appl.* **40** (2019), 870–887.
- [367] P. Ferrari et al., The eigenvalue distribution of special 2-by-2 block matrix-sequences with applications to the case of symmetrized Toeplitz structures, *SIAM J. Matrix Anal. Appl.* **40** (2019), 1066–1086.
- [368] M. Mazza and J. Pestana, Spectral properties of flipped Toeplitz matrices and related preconditioning, *BIT Numer. Math.* **59** (2019), 463–482.
- [369] J. Pestana, Nonstandard inner products and preconditioned iterative methods, PhD Thesis, University of Oxford, 2011.
- [370] A. Günnel, R. Herzog, and E. Sachs, A note on preconditioners and scalar products in Krylov subspace methods for self-adjoint problems in Hilbert space, *Electron. Trans. Numer. Anal.* **41** (2014), 13–20.
- [371] J. H. Bramble and J. E. Pasciak, A preconditioning technique for indefinite systems resulting from mixed approximations of elliptic problems, *Math. Comput.* **50** (1988), 1–17.
- [372] P. Krzyżanowski, On block preconditioners for saddle point problems with singular or indefinite (1,1) block, *Numer. Linear Algebra Appl.* **18** (2011), 123–140.
- [373] P. Concus and G. H. Golub, *A generalized conjugate gradient method for nonsymmetric systems of linear equations*, Springer, New York, NY, 1976, 56–65.
- [374] O. Widlund, A Lanczos method for a class of nonsymmetric systems of linear equations, *SIAM J. Numer. Anal.* **15** (1978), 801–812.
- [375] D. Calvetti, B. Lewis, and L. Reichel, On the regularizing properties of the GMRES method, *Numer. Math.* **91** (2002), 605–625.
- [376] M. Hanke, On Lanczos based methods for the regularization of discrete ill-posed problems, *BIT Numer. Math.* **41** (2001), 1008–1018.
- [377] T. K. Jensen and P. C. Hansen, Iterative regularization with minimum-residual methods, *BIT Numer. Math.* **47** (2007), 103–120.
- [378] M. Hanke, J. G. Nagy, and R. J. Plemmons, *Preconditioned iterative regularization for ill-posed problems*, in *Numerical Linear Algebra*, L. Reichel, A. Ruttan, and R. S. Varga, Eds., de Gruyter, Berlin, Germany, 1993, 141–164.
- [379] M. Hanke and J. Nagy, Inverse Toeplitz preconditioners for ill-posed problems, *Linear Algebra Appl.* **284** (1998), 137–156.
- [380] J. G. Nagy and K. M. Palmer, Steepest descent, CG, and iterative regularization of ill-posed problems, *BIT Numer. Math.* **43** (2003), 1003–1017.
- [381] L. Eldén and V. Simoncini, Solving ill-posed linear systems with GMRES and a singular preconditioner, *SIAM J. Matrix Anal. Appl.* **33** (2012), 1369–1394.
- [382] Z. Ranjbar and L. Eldén, Solving an ill-posed Cauchy problem for a two-dimensional parabolic PDE with variable coefficients using a preconditioned GMRES method, *SIAM J. Sci. Comput.* **36** (2014), B868–B886.
- [383] J. Baglama and L. Reichel, Decomposition methods for large linear discrete ill-posed problems, *J. Comput. Appl. Math.* **198** (2007), 332–343.
- [384] S. Gazzola and J. G. Nagy, Generalized Arnoldi–Tikhonov method for sparse reconstruction, *SIAM J. Sci. Comput.* **36** (2014), B225–B247.
- [385] P. C. Hansen and T. K. Jensen, Smoothing-norm preconditioning for regularizing minimum-residual methods, *SIAM J. Matrix Anal. Appl.* **29** (2006), 1–14.

- [386] D. Calvetti, Preconditioned iterative methods for linear discrete ill-posed problems from a Bayesian inversion perspective, *J. Comput. Appl. Math.* **198** (2007), 378–395.
- [387] V. Rokhlin and M. Tygert, A fast randomized algorithm for overdetermined linear least-squares regression, *Proc. Natl. Acad. Sci.* **105** (2008), 13212–13217.
- [388] P. Drineas et al., Faster least squares approximation, *Numer. Math.* **117** (2011), 219–249.
- [389] H. Avron, P. Maymounkov, and S. Toledo, Blendenpik: Supercharging LAPACK's least-squares solver, *SIAM J. Sci. Comput.* **32** (2010), 1217–1236.
- [390] E. S. Coakley, V. Rokhlin, and M. Tygert, A fast randomized algorithm for orthogonal projection, *SIAM J. Sci. Comput.* **33** (2011), 849–868.
- [391] X. Meng, M. A. Saunders, and M. W. Mahoney, LSRN: A parallel iterative solver for strongly over- or underdetermined systems, *SIAM J. Sci. Comput.* **36** (2014), C95–C118.
- [392] J. Yang et al., Weighted SGD for ℓ_p regression with randomized preconditioning, *J. Mach. Learn. Res.* **18** (2018), 1–43.
- [393] D. Calvetti and E. Somersalo, Priorconditioners for linear systems, *Inverse Probl.* **21** (2005), 1397–1418.
- [394] D. Calvetti, D. McGivney, and E. Somersalo, Left and right preconditioning for electrical impedance tomography with structural information, *Inverse Probl.* **28** (2012), 055015.
- [395] P. Hennig, Probabilistic interpretation of linear solvers, *SIAM J. Optim.* **25** (2015), 234–260.
- [396] J. Cockayne et al., A Bayesian conjugate gradient method (with discussion), *Bayesian Anal.* **14** (2019), 937–1012.
- [397] S. Bartels et al., Probabilistic linear solvers: A unifying view, *Stat. Comput.* **29** (2019), 1249–1263.
- [398] L. Grasedyck, Hierarchical singular value decomposition of tensors, *SIAM J. Matrix Anal. Appl.* **31** (2010), 2029–2054.
- [399] W. Hackbusch and S. Kühn, A new scheme for the tensor representation, *J. Fourier Anal. Appl.* **15** (2009), 706–722.
- [400] I. V. Oseledets, Tensor-train decomposition, *SIAM J. Sci. Comput.* **33** (2011), 2295–2317.
- [401] I. V. Oseledets, Approximation of $2^d \times 2^d$ matrices using tensor decomposition, *SIAM J. Matrix Anal. Appl.* **31** (2010), 2130–2145.
- [402] E. E. Tyrtshnikov, Tensor approximation of matrices generated by asymptotically smooth functions, *Sbornik Math.* **194** (2003), 147–160.
- [403] D. Kressner and C. Tobler, Krylov subspace methods for linear systems with tensor product structure, *SIAM J. Matrix Anal. Appl.* **31** (2010), 1688–1714.
- [404] D. Palitta and V. Simoncini, Optimality properties of Galerkin and Petrov–Galerkin methods for linear matrix equations, *Vietnam J. Math.* (2020). <https://doi.org/10.1007/s10013-020-00390-7>.
- [405] R. Andreev and C. Tobler, Multilevel preconditioning and low-rank tensor iteration for space-time simultaneous discretizations of parabolic PDEs, *Numer. Linear Alg. Appl.* **22** (2015), 317–337.
- [406] T. Breiten, V. Simoncini, and M. Stoll, Low-rank solvers for fractional differential equations, *Electron. Trans. Numer. Anal.* **45** (2016), 107–132.
- [407] S. V. Dolgov, TT-GMRES: Solution to a linear system in the structured tensor format, *Russ. J. Numer. Anal. M.* **28** (2013), 149–172.
- [408] B. N. Khoromskij, Tensor-structured preconditioners and approximate inverse of elliptic operators in \mathbb{R}^d , *Constr. Approx.* **30** (2009), 599–620.
- [409] S. Dolgov and M. Stoll, Low-rank solution to an optimization problem constrained by the Navier–Stokes equations, *SIAM J. Sci. Comput.* **39** (2017), A255–A280.
- [410] M. Stoll and T. Breiten, A low-rank in time approach to PDE-constrained optimization, *SIAM J. Sci. Comput.* **37** (2015), B1–B29.
- [411] S. Dolgov et al., Fast tensor product solvers for optimization problems with fractional differential equations as constraints, *Appl. Math. Comput.* **273** (2016), 604–623.
- [412] G. Heide, V. Khoromskaia, B. N. Khoromskij, V. Schulz, Tensor product method for fast solution of optimal control problems with fractional multidimensional Laplacian in constraints, *arXiv e-prints*. 2020: arXiv:1809.01971.
- [413] E. C. Carson et al., The numerical stability analysis of pipelined conjugate gradient methods: Historical context and methodology, *SIAM J. Sci. Comput.* **40** (2018), A3549–A3580.
- [414] J. Dongarra, M. A. Heroux, and P. Luszczek, High-performance conjugate-gradient benchmark: A new metric for ranking high-performance computing systems, *Int. J. High Perform. Comput. Appl.* **30** (2016), 3–10.
- [415] M. Hoemmen, Communication-avoiding Krylov subspace methods, PhD Thesis, University of California, Berkeley, 2010.
- [416] H. Anzt et al., *Efficiency of general Krylov methods on GPUs – An experimental study*, in *IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW)*, 2016, 683–691.
- [417] V. Eijkhout, *Introduction to high performance scientific computing*, 2nd ed., Lulu Group International, 2015.
- [418] B. Bergen et al., A massively parallel multigrid method for finite elements, *Comput. Sci. Eng.* **8** (2006), 56–62.
- [419] V. E. Henson and U. M. Yang, BoomerAMG: A parallel algebraic multigrid solver and preconditioner, *Appl. Numer. Math.* **41** (2002), 155–177.
- [420] E. Chow, Parallel implementation and practical use of sparse approximate inverse preconditioners with a priori sparsity patterns, *Int. J. High Perform. Comput. Appl.* **15** (2001), 56–74.
- [421] Janna C., Ferronato M., Sartoretto F., Gambolati G., FSAIPACK: A software package for high-performance factored sparse approximate inverse preconditioning, *ACM Trans. Math. Softw.* 2015;**41**:Art. 10.
- [422] S. Badia, A. F. Martín, and J. Principe, Multilevel balancing domain decomposition at extreme scales, *SIAM J. Sci. Comput.* **38** (2016), C22–C52.
- [423] V. A. P. Magri et al., Multilevel approaches for FSAI preconditioning, *Numer. Linear Algebra Appl.* **25** (2018), e2183.

- [424] M. Bernaschi et al., A dynamic pattern factored sparse approximate inverse preconditioner on graphics processing units, *SIAM J. Sci. Comput.* **41** (2019), C139–C160.
- [425] E. Chow and A. Patel, Fine-grained parallel incomplete LU factorization, *SIAM J. Sci. Comput.* **37** (2015), C169–C193.
- [426] H. Anzt et al., Incomplete sparse approximate inverses for parallel preconditioning, *Parallel Comput.* **71** (2018), 1–22.
- [427] H. Ibeid, R. Yokota, D. Keyes, A matrix-free preconditioner for the Helmholtz equation based on the fast multipole method, arXiv e-prints. 2016: arXiv:1608.02461.

How to cite this article: Pearson JW, Pestana J. Preconditioners for Krylov subspace methods: An overview. *GAMM-Mitteilungen*. 2020;43:e202000015. <https://doi.org/10.1002/gamm.202000015>