Electronic Transactions on Numerical Analysis.
Volume 44, pp. 522–547, 2015.

ISSN 1068–9613.

ETNA
Kent State University
http://etna.math.kent.edu

PRECONDITIONED RECYCLING KRYLOV SUBSPACE METHODS FOR SELF-ADJOINT PROBLEMS*

ANDRÉ GAUL† AND NICO SCHLÖMER‡

Abstract. A recycling Krylov subspace method for the solution of a sequence of self-adjoint linear systems is proposed. Such problems appear, for example, in the Newton process for solving nonlinear equations. Ritz vectors are automatically extracted from one MINRES run and then used for self-adjoint deflation in the next. The method is designed to work with arbitrary inner products and arbitrary self-adjoint positive-definite preconditioners whose inverse can be computed with high accuracy. Numerical experiments with nonlinear Schrödinger equations indicate a substantial decrease in computation time when recycling is used.

Key words. Krylov subspace methods, MINRES, deflation, Ritz vector recycling, nonlinear Schrödinger equations, Ginzburg–Landau equations

AMS subject classifications. 65F10, 65F08, 35Q55, 35Q56

1. Introduction.

Sequences of linear algebraic systems frequently occur in the numerical solution process of various kinds of problems. Most notable are implicit time-stepping schemes and Newton's method for solving nonlinear equation systems. It is often the case that the operators in subsequent linear systems have similar spectral properties or are in fact equal. To exploit this, a common approach is to factorize the operator once and apply the factorization to the following steps. However, this strategy typically has high memory requirements and is thus hardly applicable to problems with many unknowns. Also, it is not applicable if subsequent linear operators are even only slightly different from each other.

The authors use the idea of an alternative approach that carries over spectral information from one linear system to the next by extracting approximations of eigenvectors and using them in a deflation framework [2, 29, 30, 41]. For a more detailed overview on the background of such methods, see [14]. The method is designed for Krylov subspace methods in general and is worked out in this paper for the MINRES method [37] in particular.

The idea of recycling spectral information in Krylov subspace methods is not new. Notably, Kilmer and de Sturler [24] adapted the GCRO method [3] for recycling in the setting of a sequence of linear systems. Essentially, this strategy consists of applying the MINRES method to a projected linear system where the projection is built from approximate eigenvectors for the first matrix of the sequence. Wang, de Sturler, and Paulino proposed the RMINRES method [52] that also includes the extraction of approximate eigenvectors. In contrast to Kilmer and de Sturler, the RMINRES method is a modification of the MINRES method that explicitly includes these vectors in the search space for the following linear systems (*augmentation*). Similar recycling techniques based on GCRO have also been used by Parks et al. [38], Mello et al. [28], Feng, Benner, and Korvink [12], and Soodhalter, Szyld, and Xue [49]. A different approach has been proposed by Giraud, Gratton, and Martin [19], where a preconditioner is updated with approximate spectral information for use in a GMRES variant.

GCRO-based methods with augmentation of the search space, including RMINRES, are mathematically equivalent to the standard GMRES method (or MINRES for the symmetric

*Received September 19, 2013. Accepted May 19, 2015. Recommended by D. Szyld. Published online on October 7, 2015. The work of André Gaul was supported by the DFG Forschungszentrum MATHEON. The work of Nico Schlömer was supported by the Research Foundation Flanders (FWO).

†Institut für Mathematik, Technische Universität Berlin, Straße des 17. Juni, D-10623 Berlin, Germany (gaul@math.tu-berlin.de).

‡Departement Wiskunde en Informatica, Universiteit Antwerpen, Middelheimlaan 1, B-2020 Antwerpen, Belgium (nico.schloemer@ua.ac.be).

ETNA
Kent State University
http://etna.math.kent.edu

case) applied to a projected linear system [14]. Krylov subspace methods that are applied to projected linear systems are often called *deflated* methods. In the literature, both augmented and deflated methods have been used in a variety of settings; we refer to Eiermann, Ernst, and Schneider [9] and the review article by Simoncini and Szyld [45] for a comprehensive overview.

In general, Krylov subspace methods are only feasible in combination with a preconditioner. In [52] only factorized preconditioners of the form $A \approx CC^T$ can be used such that instead of $Ax = b$ the preconditioned system $C^{-1}AC^{-T}y = C^{-1}b$ is solved. In this case the system matrix remains symmetric. While preconditioners like (incomplete) Cholesky factorizations have this form, other important classes like (algebraic) multigrid do not. A major difference of the method presented here from RMINRES is that it allows for a greater variety of preconditioners. The only restrictions on the preconditioner M^{-1} are that it has to be self-adjoint and positive-definite and that its inverse has to be known; see the discussion at the end of Section 2.3 for more details. While this excludes the popular class of multigrid preconditioners with a fixed number of cycles, full multigrid preconditioners are admissible. To the best knowledge of the authors, no such method has been considered before. Note that the requirement of a self-adjoint and positive-definite preconditioner M^{-1} is common in the context of methods for self-adjoint problems (e.g., CG and MINRES) because it allows to change the inner product implicitly. With such a preconditioner, the inertia of A is preserved in $M^{-1}A$. Deflation is able to remedy the problem to a certain extent, e.g., if A has only a few negative eigenvalues.

Moreover, general inner products are considered, which facilitate the incorporation of arbitrary preconditioners and allow to exploit the self-adjointness of a problem algorithmically when its natural inner product is used. This leads to an efficient three-term recurrence with the MINRES method instead of a costly full orthogonalization in GMRES. One important example of problems that are self-adjoint with respect to a non-Euclidean inner product are nonlinear Schrödinger equations, presented in more detail in Section 3. General inner products have been considered before; see, e.g., Freund, Golub, and Nachtigal [13] or Eiermann, Ernst, and Schneider [9]. Naturally, problems which are Hermitian (with respect to the Euclidean inner product) also benefit from the results in this work.

In many of the previous approaches, the underlying Krylov subspace method itself has to be modified for including deflation; see, e.g., the *modified MINRES method* of Wang, de Sturler, and Paulino [52, Algorithm 1]. In contrast, the work in the present paper separates the deflation methodology from the Krylov subspace method. Deflation can thus be implemented algorithmically as a wrapper around any existing MINRES code, e.g., by optimized high-performance variants. The notes on the implementation in Sections 2.2 and 2.3 discuss efficient realizations thereof.

For the sake of clarity, restarting—often used to mitigate memory constraints—is not explicitly discussed in the present paper. However, as noted in Section 2.3, it can be added easily to the algorithm without affecting the presented advantages of the method. Note that the method in [52] does not require restarting because it computes Ritz vectors from a fixed number of Lanczos vectors (*cycling*); cf. Section 2.3. Since the non-restarted method maintains global optimality over the entire Krylov subspace (in exact arithmetic), it may exhibit a more favorable convergence behavior than restarted methods.

In addition to the deflation of computed Ritz vectors, other vectors can be included that carry explicit information about the problem in question. For example, approximations to eigenvectors corresponding to critical eigenvalues are readily available from analytic considerations. Applications for this are plentiful, e.g., flow in porous media considered by Tang et al. [51] and nonlinear Schrödinger equations; see Section 3.

ETNA
Kent State University
http://etna.math.kent.edu

The deflation framework with the properties presented in this paper are applied in the numerical solution of nonlinear Schrödinger equations. Nonlinear Schrödinger equations and their variations are used to describe a wide variety of physical systems, most notably in superconductivity, quantum condensates, nonlinear acoustics [48], nonlinear optics [17], and hydrodynamics [34]. For the solution of nonlinear Schrödinger equations with Newton's method, a linear system has to be solved with the Jacobian operator for each Newton update. The Jacobian operator is self-adjoint with respect to a non-Euclidean inner product and indefinite. In order to be applicable in practice, the MINRES method can be combined with an AMG-type preconditioner that is able to limit the number of MINRES iterations to a feasible extent [43]. Due to the special structure of the nonlinear Schrödinger equation, the Jacobian operator exhibits one eigenvalue that moves to zero when the Newton iterate converges to a nontrivial solution and is exactly zero at a solution. Because this situation only occurs in the last step, no linear system has to be solved with an exactly singular Jacobian operator but the severely ill-conditioned Jacobian operators in the final Newton steps lead to convergence slowdown or stagnation in the MINRES method even when a preconditioner is applied. For the numerical experiments, we consider the Ginzburg–Landau equation, an important instance of nonlinear Schrödinger equations, that models phenomena of certain superconductors. We use the proposed recycling MINRES method and show how it can help to improve the convergence of the MINRES method. All enhancements of the deflated MINRES method, i.e., arbitrary inner products and preconditioners, are required for this application. As a result, the overall time consumption of Newton's method for the Ginzburg–Landau equation is reduced by roughly 40%.

The deflated Krylov subspace methods described in this paper are implemented in the Python package *KryPy* [15]; solvers for nonlinear Schrödinger problems are available from *PyNosh* [16]. Both packages are free and open-source software. All results from this paper can be reproduced with the help of these packages.

The paper is organized as follows: Section 2 gives a brief overview on the preconditioned MINRES method for an arbitrary nonsingular linear operator that is self-adjoint with respect to an arbitrary inner product. The deflated MINRES method is described in Section 2.2 while Section 2.3 presents the computation of Ritz vectors and explains their use in the overall algorithm for the solution of a sequence of self-adjoint linear systems. In Section 3 this algorithm is applied to the Ginzburg–Landau equation. Sections 3.1 and 3.2 deal with the numerical treatment of nonlinear Schrödinger equations in general and the Ginzburg–Landau equation in particular. In Section 3.3 numerical results for typical two- and three-dimensional setups are presented.

2. The MINRES method.

2.1. Preconditioned MINRES with arbitrary inner product. This section presents well-known properties of the preconditioned MINRES method. As opposed to ordinary textbook presentations, this section incorporates a general Hilbert space. For $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$ let H be a $\mathbb{K}$-Hilbert space with inner product $\langle \cdot, \cdot \rangle_H$ and induced norm $\|\cdot\|_H$. Throughout this paper, the inner product $\langle \cdot, \cdot \rangle_H$ is linear in the first and anti-linear in the second argument and we define $L(H) := \{\mathcal{L} : H \longrightarrow H \mid \mathcal{L} \text{ is linear and bounded}\}$. The vector space of k-by-l matrices is denoted by $\mathbb{K}^{k,l}$. We wish to obtain $x \in H$ from

$$\mathcal{A}x = b, \tag{2.1}$$

where $\mathcal{A} \in L(H)$ is $\langle \cdot, \cdot \rangle_H$-self-adjoint and invertible and $b \in H$. The self-adjointness implies that the spectrum $\sigma(\mathcal{A})$ is real. However, we do not assume that $\mathcal{A}$ is definite.

ETNA
Kent State University
http://etna.math.kent.edu

Given an initial guess $x_0 \in H$, we can approximate x by the iterates

$$x_n = x_0 + y_n \quad \text{with} \quad y_n \in \mathcal{K}_n(\mathcal{A}, r_0), \tag{2.2}$$

where $r_0 = b - \mathcal{A}x_0$ is the initial residual and $\mathcal{K}_n(\mathcal{A}, r_0) = \operatorname{span}\{r_0, \mathcal{A}r_0, \ldots, \mathcal{A}^{n-1}r_0\}$ is the nth Krylov subspace generated by $\mathcal{A}$ and r_0. We concentrate on minimal residual methods here, i.e., methods that construct iterates of the form (2.2) such that the residual $r_n := b - \mathcal{A}x_n$ has minimal $\|\cdot\|_H$-norm, that is,

$$\begin{aligned} \|r_n\|_H &= \|b - \mathcal{A}x_n\|_H = \|b - \mathcal{A}(x_0 + y_n)\|_H = \|r_0 - \mathcal{A}y_n\|_H \\ &= \min_{y \in \mathcal{K}_n(\mathcal{A}, r_0)} \|r_0 - \mathcal{A}y\|_H = \min_{p \in \Pi_n^0} \|p(\mathcal{A})r_0\|_H , \end{aligned} \tag{2.3}$$

where Π_n^0 is the set of polynomials of degree at most n with $p(0) = 1$. For a general invertible linear operator $\mathcal{A}$, the minimization problem in (2.3) can be solved by the GMRES method [40] which is mathematically equivalent to the MINRES method [37] if $\mathcal{A}$ is self-adjoint [26, Section 2.5.5].

To facilitate subsequent definitions and statements for general Hilbert spaces, we use a block notation for inner products that generalizes the common block notation for matrices:

DEFINITION 2.1. *For $k, l \in \mathbb{N}$ and two tuples of vectors $X = [x_1, \ldots, x_k] \in H^k$ and $Y = [y_1, \ldots, y_l] \in H^l$, the product $\langle \cdot, \cdot \rangle_H : H^k \times H^l \longrightarrow \mathbb{K}^{k,l}$ is defined by*

$$\langle X, Y \rangle_H := \left[\langle x_i, y_j \rangle_H \right]_{\substack{i=1,\ldots,k \\ j=1,\ldots,l}} .$$

For the Euclidean inner product and two matrices $X \in \mathbb{C}^{N,k}$ and $Y \in \mathbb{C}^{N,l}$, the product takes the form $\langle X, Y \rangle_2 = X^{\mathrm{H}} Y$.

A block $X \in H^k$ can be right-multiplied with a matrix just as in the plain matrix case:

DEFINITION 2.2. *For $X \in H^k$ and $\boldsymbol{Z} = [z_{ij}]_{\substack{i=1,\ldots,k \\ j=1,\ldots,l}} \in \mathbb{K}^{k,l}$, right multiplication is defined by*

$$X\boldsymbol{Z} := \left[\sum_{i=1}^{k} z_{ij} x_i \right]_{j=1,\ldots,l} \in H^l .$$

Because the MINRES method and the underlying Lanczos algorithm are often described for Hermitian matrices only (i.e., for the Euclidean inner product), we recall very briefly some properties of the Lanczos algorithm for a linear operator that is self-adjoint with respect to an arbitrary inner product $\langle \cdot, \cdot \rangle_H$ [8]. If the Lanczos algorithm with inner product $\langle \cdot, \cdot \rangle_H$ applied to $\mathcal{A}$ and the initial vector $v_1 = r_0 / \|r_0\|_H$ has completed the nth iteration, then the Lanczos relation

$$\mathcal{A}V_n = V_{n+1}\underline{\boldsymbol{T}}_n$$

holds, where the elements of $V_{n+1} = [v_1, \ldots, v_{n+1}] \in H^{n+1}$ form a $\langle \cdot, \cdot \rangle_H$-orthonormal basis of $\mathcal{K}_{n+1}(\mathcal{A}, r_0)$, i.e., $\operatorname{span}\{v_1, \ldots, v_{n+1}\} = \mathcal{K}_{n+1}(\mathcal{A}, r_0)$ and $\langle V_{n+1}, V_{n+1} \rangle_H = \boldsymbol{I}_{n+1}$. Note that the orthonormality implies $\|V_{n+1}z\|_H = \|z\|_2$ for all $z \in \mathbb{K}^{n+1}$. The matrix $\underline{\boldsymbol{T}}_n \in \mathbb{R}^{n+1,n}$ is real-valued (even if $\mathbb{K} = \mathbb{C}$), symmetric, and tridiagonal with

$$\underline{\boldsymbol{T}}_k = [\langle \mathcal{A}v_i, v_j \rangle_H]_{\substack{i=1,\ldots,n+1 \\ j=1,\ldots,n}} .$$

The nth approximation of the solution of the linear system (2.1) generated with the MINRES method and the corresponding residual norm, cf. (2.2) and (2.3), can then be expressed as

$$\begin{aligned} x_n &= x_0 + V_n z_n \quad \text{with} \quad z_n \in \mathbb{K}^n \quad \text{and} \\ \|r_n\|_H &= \|r_0 - \mathcal{A}V_n z_n\|_H = \|V_{n+1}(\|r_0\|_H e_1 - \underline{\boldsymbol{T}}_n z_n)\|_H = \|\|r_0\|_H e_1 - \underline{\boldsymbol{T}}_n z_n\|_2 . \end{aligned}$$

ETNA
Kent State University
http://etna.math.kent.edu

By recursively computing a QR decomposition of $\underline{\boldsymbol{T}}_n$, the minimization problem in (2.3) can be solved without storing the entire matrix $\underline{\boldsymbol{T}}_n$ and, more importantly, the full Lanczos basis V_n.

Let $N := \dim H < \infty$, and let the elements of $W \in H^N$ form a $\langle \cdot, \cdot \rangle_H$-orthonormal basis of H consisting of eigenvectors of $\mathcal{A}$. Then $\mathcal{A}W = W\boldsymbol{D}$ for the diagonal matrix $\boldsymbol{D} = \operatorname{diag}(\lambda_1, \dots, \lambda_N)$ with $\mathcal{A}$'s eigenvalues $\lambda_1, \dots, \lambda_N \in \mathbb{R}$ on the diagonal. Let $r_0^W \in \mathbb{K}^N$ be the representation of r_0 in the basis W, i.e., $r_0 = W r_0^W$. According to (2.3), the residual norm of the nth approximation obtained with MINRES can be expressed as

$$\begin{aligned} \|r_n\|_H &= \min_{p \in \Pi_n^0} \left\| p(\mathcal{A}) W r_0^W \right\|_H = \min_{p \in \Pi_n^0} \left\| W p(\boldsymbol{D}) r_0^W \right\|_H = \min_{p \in \Pi_n^0} \left\| p(\boldsymbol{D}) r_0^W \right\|_2 \\ &\leq \left\| r_0^W \right\|_2 \min_{p \in \Pi_n^0} \|p(\boldsymbol{D})\|_2 \,. \end{aligned} \tag{2.4}$$

From $\left\| r_0^W \right\|_2 = \left\| W r_0^W \right\|_H = \|r_0\|_H$ and $\|p(\boldsymbol{D})\|_2 = \max_{i \in \{1,\dots,N\}} |p(\lambda_i)|$, we obtain the well-known MINRES worst-case bound for the relative residual norm [20, 27]

$$\frac{\|r_n\|_H}{\|r_0\|_H} \leq \min_{p \in \Pi_n^0} \max_{i \in \{1,\dots,N\}} |p(\lambda_i)|. \tag{2.5}$$

This can be estimated even further upon letting the eigenvalues of $\mathcal{A}$ be sorted such that $\lambda_1 \leq \cdots \leq \lambda_s < 0 < \lambda_{s+1} \leq \cdots \leq \lambda_N$ for a $s \in \mathbb{N}_0$. By replacing the discrete set of eigenvalues in (2.5) by the union of the two intervals $I^- := [\lambda_1, \lambda_s]$ and $I^+ := [\lambda_{s+1}, \lambda_N]$, one gets

$$\begin{aligned} \frac{\|r_n\|_H}{\|r_0\|_H} &\leq \min_{p \in \Pi_n^0} \max_{\lambda \in \sigma(\mathcal{A})} |p(\lambda)| \leq \min_{p \in \Pi_n^0} \max_{\lambda \in I^- \cup I^+} |p(\lambda)| \\ &\leq 2 \left(\frac{\sqrt{|\lambda_1 \lambda_N|} - \sqrt{|\lambda_s \lambda_{s+1}|}}{\sqrt{|\lambda_1 \lambda_N|} + \sqrt{|\lambda_s \lambda_{s+1}|}} \right)^{[n/2]}, \end{aligned} \tag{2.6}$$

where $[n/2]$ is the integer part of $n/2$; cf. [20, 27]. Note that this estimate does not take into account the actual distribution of the eigenvalues in the intervals I^- and I^+. In practice a better convergence behavior than the one suggested by the estimate above can often be observed.

In most applications, the MINRES method is only feasible when it is applied with a preconditioner. Consider the preconditioned system

$$\mathcal{M}^{-1} \mathcal{A} x = \mathcal{M}^{-1} b, \tag{2.7}$$

where $\mathcal{M} \in L(H)$ is a $\langle \cdot, \cdot \rangle_H$-self-adjoint, invertible, and positive-definite linear operator. Note that $\mathcal{M}^{-1}\mathcal{A}$ is not $\langle \cdot, \cdot \rangle_H$-self-adjoint but self-adjoint with respect to the inner product $\langle \cdot, \cdot \rangle_{\mathcal{M}}$ defined by $\langle x, y \rangle_{\mathcal{M}} := \langle \mathcal{M}x, y \rangle_H = \langle x, \mathcal{M}y \rangle_H$. The MINRES method is then applied to (2.7) with the inner product $\langle \cdot, \cdot \rangle_{\mathcal{M}}$ and thus minimizes $\left\| \mathcal{M}^{-1}(b - \mathcal{A}x) \right\|_{\mathcal{M}}$. From an algorithmic point of view it is worthwhile to note that only the application of $\mathcal{M}^{-1}$ is needed and the application of $\mathcal{M}$ for the inner products can be carried out implicitly; cf. [11, Chapter 6]. Analogously to (2.6) the convergence bound for the residuals $\widetilde{r}_n$ produced by the preconditioned MINRES method is

$$\frac{\|\widetilde{r}_n\|_{\mathcal{M}}}{\|\mathcal{M}^{-1} r_0\|_{\mathcal{M}}} \leq \min_{p \in \Pi_n^0} \max_{\mu \in \sigma(\mathcal{M}^{-1}\mathcal{A})} |p(\mu)| \,.$$

Thus the goal of preconditioning is to achieve a more favorable spectrum of $\mathcal{M}^{-1}\mathcal{A}$ with an appropriate $\mathcal{M}^{-1}$.

ETNA
Kent State University
http://etna.math.kent.edu

2.2. Deflated MINRES. In many applications even with the aid of a preconditioner, the convergence of MINRES is hampered—often due to the presence of one or a few eigenvalues close to zero that are isolated from the remaining spectrum. This case has recently been studied by Simoncini and Szyld [46]. Their analysis and numerical experiments show that an isolated simple eigenvalue can cause stagnation of the residual norm until a harmonic Ritz value approximates the outlying eigenvalue well.

Two strategies are well-known in the literature to circumvent the stagnation or slowdown in the convergence of preconditioned Krylov subspace methods described above: *augmentation* and *deflation*. In augmented methods, the Krylov subspace is enlarged by a suitable subspace that contains useful information about the problem. In deflation techniques, the operator is modified with a suitably chosen projection in order to "eliminate" components that hamper convergence; e.g., eigenvalues close to the origin. For an extensive overview of these techniques we refer to Eiermann, Ernst, and Schneider [9] and the survey article by Simoncini and Szyld [45]. Both techniques are closely intertwined and even turn out to be equivalent in some cases [14]. Here, we concentrate on deflated methods and first give a brief description of the recycling MINRES (RMINRES) method introduced by Wang, de Sturler, and Paulino [52] before presenting a slightly different approach.

The RMINRES method by Wang, de Sturler, and Paulino [52] is mathematically equivalent [14] to the application of the MINRES method to the "deflated" equation

$$\mathcal{P}_1\mathcal{A}\widetilde{x} = \mathcal{P}_1 b, \tag{2.8}$$

where for a given d-tuple $U \in H^d$ of linearly independent vectors (which constitute a basis of the recycling space) and $C := \mathcal{A}U$, the linear operator $\mathcal{P}_1 \in L(H)$ is defined by $\mathcal{P}_1 x := x - C\langle C, C\rangle_H^{-1}\langle C, x\rangle_H$. Note that, although $\mathcal{P}_1$ is a $\langle\cdot,\cdot\rangle_H$-self-adjoint projection (and thus an orthogonal projection), $\mathcal{P}_1\mathcal{A}$ in general is not. However, as outlined in [52, Section 4], an orthonormal basis of the Krylov subspace can still be generated with MINRES' short recurrences and the operator $\mathcal{P}_1\mathcal{A}$ because $\mathcal{K}_n(\mathcal{P}_1\mathcal{A}, \mathcal{P}_1 r_0) = \mathcal{K}_n(\mathcal{P}_1\mathcal{A}\mathcal{P}_1^*, \mathcal{P}_1 r_0)$. Solutions of equation (2.8) are not unique for $d > 0$ and thus x was replaced by $\widetilde{x}$. To obtain an approximation x_n of the original solution x from the approximation $\widetilde{x}_n$ generated with MINRES applied to (2.8), an additional correction has to be applied:

$$x_n = \widetilde{\mathcal{P}}_1\widetilde{x}_n + U\langle C, C\rangle_H^{-1}\langle C, b\rangle_H,$$

where $\widetilde{\mathcal{P}}_1 \in L(H)$ is defined by $\widetilde{\mathcal{P}}_1 x := x - U\langle C, C\rangle_H^{-1}\langle C, \mathcal{A}x\rangle_H$.

Let us now turn to a slightly different deflation technique for MINRES which we formulate with preconditioning directly. We will use a projection which has been developed in the context of the CG method for Hermitian and positive-definite operators [4, 33, 51]. Under a mild assumption, this projection is also well-defined in the indefinite case. In contrast to the orthogonal projection $\mathcal{P}_1$ used in RMINRES, it is not self-adjoint but instead renders the projected operator self-adjoint. This is a natural fit for an integration with the MINRES method.

Our goal is to use approximations to eigenvectors corresponding to eigenvalues that hamper convergence in order to modify the operator with a projection. Consider the preconditioned equation (2.7) and assume for a moment that the elements of $U = [u_1, \ldots, u_d] \in H^d$ form a $\langle\cdot,\cdot\rangle_{\mathcal{M}}$-orthonormal basis consisting of eigenvectors of $\mathcal{M}^{-1}\mathcal{A}$, i.e., $\mathcal{M}^{-1}\mathcal{A}U = U\boldsymbol{D}$ with a diagonal matrix $\boldsymbol{D} = \operatorname{diag}(\lambda_1, \ldots, \lambda_d) \in \mathbb{R}^{d,d}$. Then $\langle U, \mathcal{M}^{-1}\mathcal{A}U\rangle_{\mathcal{M}} = \langle U, U\rangle_{\mathcal{M}}\boldsymbol{D} = \boldsymbol{D}$ is nonsingular because we assumed that $\mathcal{A}$ is invertible. This motivates the following definition:

ETNA
Kent State University
http://etna.math.kent.edu

DEFINITION 2.3. *Let $\mathcal{M}, \mathcal{A} \in L(H)$ be invertible and $\langle \cdot, \cdot \rangle_H$-self-adjoint operators, and let $\mathcal{M}$ be positive-definite. Let $U \in H^d$ be such that $\langle U, \mathcal{M}^{-1}\mathcal{A}U \rangle_{\mathcal{M}} = \langle U, \mathcal{A}U \rangle_H$ is nonsingular. We define the projections $\mathcal{P}_{\mathcal{M}}, \mathcal{P} \in L(H)$ by*

$$\begin{aligned} \mathcal{P}_{\mathcal{M}} x &:= x - \mathcal{M}^{-1}\mathcal{A}U \langle U, \mathcal{M}^{-1}\mathcal{A}U \rangle_{\mathcal{M}}^{-1} \langle U, x \rangle_{\mathcal{M}} \qquad \text{and} \\ \mathcal{P} x &:= x - \mathcal{A}U \langle U, \mathcal{A}U \rangle_H^{-1} \langle U, x \rangle_H. \end{aligned} \tag{2.9}$$

The projection $\mathcal{P}_{\mathcal{M}}$ is the projection onto $\operatorname{range}(U)^{\perp_{\mathcal{M}}}$ along $\operatorname{range}(\mathcal{M}^{-1}\mathcal{A}U)$, whereas $\mathcal{P}$ is the projection onto $\operatorname{range}(U)^{\perp_H}$ along $\operatorname{range}(\mathcal{A}U)$.

The assumption in Definition 2.3 that $\langle U, \mathcal{M}^{-1}\mathcal{A}U \rangle_{\mathcal{M}}$ is nonsingular holds if and only if $\operatorname{range}(\mathcal{M}^{-1}\mathcal{A}U) \cap \operatorname{range}(U)^{\perp_{\mathcal{M}}} = \{0\}$ or equivalently if $\operatorname{range}(\mathcal{A}U) \cap \operatorname{range}(U)^{\perp_H} = \{0\}$. As stated above, this condition is fulfilled if U contains a basis of eigenvectors of $\mathcal{M}^{-1}\mathcal{A}$ and also holds for good-enough approximations thereof; see, e.g., the monograph of Stewart and Sun [50] for a thorough analysis of perturbations of invariant subspaces. Applying the projection $\mathcal{P}_{\mathcal{M}}$ to the preconditioned equation (2.7) yields the deflated equation

$$\mathcal{P}_{\mathcal{M}} \mathcal{M}^{-1} \mathcal{A} \widetilde{x} = \mathcal{P}_{\mathcal{M}} \mathcal{M}^{-1} b. \tag{2.10}$$

The following lemma states some important properties of the operator $\mathcal{P}_{\mathcal{M}}\mathcal{M}^{-1}\mathcal{A}$.

LEMMA 2.4. *Let the assumptions in Definition 2.3 hold. Then*

1. $\mathcal{P}_{\mathcal{M}}\mathcal{M}^{-1} = \mathcal{M}^{-1}\mathcal{P}$.
2. $\mathcal{P}\mathcal{A} = \mathcal{A}\mathcal{P}^*$ *where* $\mathcal{P}^*$ *is the adjoint operator of* $\mathcal{P}$ *with respect to* $\langle \cdot, \cdot \rangle_H$, *defined by* $\mathcal{P}^* x = x - U \langle U, \mathcal{A}U \rangle_H^{-1} \langle \mathcal{A}U, x \rangle_H$.
3. $\mathcal{P}_{\mathcal{M}}\mathcal{M}^{-1}\mathcal{A} = \mathcal{M}^{-1}\mathcal{P}\mathcal{A} = \mathcal{M}^{-1}\mathcal{A}\mathcal{P}^*$ *is self-adjoint with respect to* $\langle \cdot, \cdot \rangle_{\mathcal{M}}$.
4. *For each initial guess* $\widetilde{x}_0 \in H$, *the MINRES method with inner product* $\langle \cdot, \cdot \rangle_{\mathcal{M}}$ *applied to equation (2.10) is well defined at each iteration until it terminates with a solution of the system.*
5. *If* $\widetilde{x}_n$ *is the* n*th approximation and* $\mathcal{P}_{\mathcal{M}}\mathcal{M}^{-1}b - \mathcal{P}_{\mathcal{M}}\mathcal{M}^{-1}\mathcal{A}\widetilde{x}_n$ *the corresponding residual generated by the MINRES method with inner product* $\langle \cdot, \cdot \rangle_{\mathcal{M}}$ *applied to (2.10) with initial guess* $\widetilde{x}_0 \in H$, *then the corrected approximation*

$$x_n := \mathcal{P}^* \widetilde{x}_n + U \langle U, \mathcal{A}U \rangle_H^{-1} \langle U, b \rangle_H \tag{2.11}$$

fulfills

$$\mathcal{M}^{-1}b - \mathcal{M}^{-1}\mathcal{A}x_n = \mathcal{P}_{\mathcal{M}}\mathcal{M}^{-1}b - \mathcal{P}_{\mathcal{M}}\mathcal{M}^{-1}\mathcal{A}\widetilde{x}_n. \tag{2.12}$$

(Note that (2.12) also holds for $n = 0$.)

Proof. Statements 1, 2, and the equation in 3 follow from elementary calculations. Because

$$\begin{aligned} \langle \mathcal{P}_{\mathcal{M}}\mathcal{M}^{-1}\mathcal{A}x, y \rangle_{\mathcal{M}} &= \langle \mathcal{P}\mathcal{A}x, y \rangle_H = \langle \mathcal{A}x, \mathcal{P}^* y \rangle_H = \langle x, \mathcal{A}\mathcal{P}^* y \rangle_H = \langle x, \mathcal{P}\mathcal{A}y \rangle_H \\ &= \langle x, \mathcal{P}_{\mathcal{M}}\mathcal{M}^{-1}\mathcal{A}y \rangle_{\mathcal{M}} \end{aligned}$$

holds for all $x, y \in H$, the operator $\mathcal{P}_{\mathcal{M}}\mathcal{M}^{-1}\mathcal{A}$ is self-adjoint with respect to $\langle \cdot, \cdot \rangle_{\mathcal{M}}$.

Statement 4 immediately follows from [14, Theorem 5.1] and the self-adjointness of $\mathcal{P}_{\mathcal{M}}\mathcal{M}^{-1}\mathcal{A}$. Note that the referenced theorem is stated for the Euclidean inner product but it can easily be generalized to arbitrary inner products. Moreover, GMRES is mathematically equivalent to MINRES in our case, again due to the self-adjointness.

Statement 5 follows from 1 and 3 by direct calculations:

$$\begin{aligned} \mathcal{M}^{-1}b - \mathcal{M}^{-1}\mathcal{A}x_n &= \mathcal{M}^{-1}(b - \mathcal{A}U \langle U, \mathcal{A}U \rangle_H^{-1} \langle U, b \rangle_H) - \mathcal{M}^{-1}\mathcal{A}\mathcal{P}^*\widetilde{x}_n \\ &= \mathcal{M}^{-1}\mathcal{P}b - \mathcal{P}_{\mathcal{M}}\mathcal{M}^{-1}\mathcal{A}\widetilde{x}_n = \mathcal{P}_{\mathcal{M}}\mathcal{M}^{-1}b - \mathcal{P}_{\mathcal{M}}\mathcal{M}^{-1}\mathcal{A}\widetilde{x}_n. \end{aligned}$$

∎

ETNA
Kent State University
http://etna.math.kent.edu

Now that we know that MINRES is well-defined when applied to the deflated and preconditioned equation (2.10), we want to investigate the convergence behavior in comparison with the original preconditioned equation (2.7). The following result is well-known for the positive-definite case; see, e.g., Saad, Yeung, Erhel, and Guyomarc'h [41]. The proof is quite canonical and given here for convenience of the reader.

LEMMA 2.5. *Let the assumptions in Definition 2.3 and* $N := \dim H < \infty$ *hold. If the spectrum of the preconditioned operator* $\mathcal{M}^{-1}\mathcal{A}$ *is* $\sigma(\mathcal{M}^{-1}\mathcal{A}) = \{\lambda_1, \ldots, \lambda_N\}$ *and for* $d > 0$ *the elements of* $U \in H^d$ *form a basis of the* $\mathcal{M}^{-1}\mathcal{A}$*-invariant subspace corresponding to the eigenvalues* $\lambda_1, \ldots, \lambda_d$*, then the following holds:*

1. *The spectrum of the deflated operator* $\mathcal{P}_{\mathcal{M}}\mathcal{M}^{-1}\mathcal{A}$ *is*

$$\sigma(\mathcal{P}_{\mathcal{M}}\mathcal{M}^{-1}\mathcal{A}) = \{0\} \cup \{\lambda_{d+1}, \ldots, \lambda_N\}.$$

2. *For* $n \geq 0$*, let* x_n *be the* n*th corrected approximation (cf. statement 5 of Lemma 2.4) of MINRES applied to (2.10) with inner product* $\langle \cdot, \cdot \rangle_{\mathcal{M}}$ *and initial guess* $\widetilde{x}_0$*. The residuals* $r_n := \mathcal{M}^{-1}b - \mathcal{M}^{-1}\mathcal{A}x_n$ *then fulfill*

$$\frac{\|r_n\|_{\mathcal{M}}}{\|r_0\|_{\mathcal{M}}} \leq \min_{p \in \Pi_n^0} \max_{i \in \{d+1, \ldots, N\}} |p(\lambda_i)|.$$

Proof. From the definition of $\mathcal{P}_{\mathcal{M}}$ in Definition 2.3, we obtain $\mathcal{P}_{\mathcal{M}}\mathcal{M}^{-1}\mathcal{A}U = 0$ and thus know that 0 is an eigenvalue of $\mathcal{P}_{\mathcal{M}}\mathcal{M}^{-1}\mathcal{A}$ with multiplicity at least d. Let the elements of $V \in H^{N-d}$ be orthonormal and such that $\mathcal{M}^{-1}\mathcal{A}V = V\boldsymbol{D}_2$ with $\boldsymbol{D}_2 = \operatorname{diag}(\lambda_{d+1}, \ldots, \lambda_N)$. Then $\langle U, V \rangle_{\mathcal{M}} = 0$ because $\mathcal{M}^{-1}\mathcal{A}$ is self-adjoint with respect to $\langle \cdot, \cdot \rangle_{\mathcal{M}}$. Thus $\mathcal{P}_{\mathcal{M}}V = V$ and the statement follows from $\mathcal{P}_{\mathcal{M}}\mathcal{M}^{-1}\mathcal{A}V = V\boldsymbol{D}_2$.

Because the residual corresponding to the corrected initial guess is

$$r_0 = \mathcal{P}_{\mathcal{M}}\mathcal{M}^{-1}(b - \mathcal{A}\widetilde{x}_0) \in \operatorname{range}(U)^{\perp_{\mathcal{M}}} = \operatorname{range}(V),$$

where V is defined as above, we have $r_0 = Vr_0^V$ for a $r_0^V \in \mathbb{K}^{N-d}$. Then with $\boldsymbol{D}_2$ as above, we obtain by using the orthonormality of V similar to (2.4):

$$\begin{aligned}\|r_n\|_{\mathcal{M}} &= \min_{p \in \Pi_n^0} \left\|p(\mathcal{P}_{\mathcal{M}}\mathcal{M}^{-1}\mathcal{A})Vr_0^V\right\|_{\mathcal{M}} = \min_{p \in \Pi_n^0} \left\|Vp(\boldsymbol{D}_2)r_0^V\right\|_{\mathcal{M}} \\ &= \min_{p \in \Pi_n^0} \left\|p(\boldsymbol{D}_2)r_0^V\right\|_2 \leq \|r_0\|_{\mathcal{M}} \min_{p \in \Pi_n^0} \max_{i \in \{d+1, \ldots, N\}} |p(\lambda_i)|. \qquad \square\end{aligned}$$

2.2.1. Notes on the implementation. By item 1 in Lemma 2.4, $\mathcal{P}_{\mathcal{M}}\mathcal{M}^{-1} = \mathcal{M}^{-1}\mathcal{P}$, the MINRES method can be applied to the linear system

$$\mathcal{M}^{-1}\mathcal{P}\mathcal{A}\widetilde{x} = \mathcal{M}^{-1}\mathcal{P}b \tag{2.13}$$

instead of (2.10). When an approximate solution $\widetilde{x}_n$ of (2.13) is satisfactory, then the correction (2.11) has to be applied to obtain an approximate solution of the original system (2.1). Note that neither $\mathcal{M}$ nor its inverse $\mathcal{M}^{-1}$ show up in the definition of the operator $\mathcal{P}$ or its adjoint operator $\mathcal{P}^*$ which is used in the correction. Thus the preconditioner $\mathcal{M}^{-1}$ does not have to be applied to additional vectors if deflation is used. This can be a major advantage since the application of the preconditioner operator $\mathcal{M}^{-1}$ is the most expensive part in many applications.

The deflation operator $\mathcal{P}$ as defined in Definition 2.3 with $U \in H^d$ needs to store $2d$ vectors because aside from U also $C := \mathcal{A}U$ should be precomputed and stored. Furthermore, the matrix $\boldsymbol{E} := \langle U, C \rangle_H \in \mathbb{K}^{d,d}$ or its inverse have to be stored. The adjoint operator $\mathcal{P}^*$

ETNA
Kent State University
http://etna.math.kent.edu

TABLE 2.1
Storage requirements and computational cost of the projection operators $\mathcal{P}$ and $\mathcal{P}^*$ (cf. Definition 2.3 and Lemma 2.4). All vectors are of length N, i.e., the number of degrees of freedom of the underlying problem. Typically, $N \gg d$.

(a) Storage requirements

	vectors	other
U	d	–
$C = \mathcal{A}U$	d	–
$\boldsymbol{E} = \langle U, C\rangle_H$ or $\boldsymbol{E}^{-1}$	–	d^2

(b) Computational cost

	applications of $\mathcal{A}$	applications of $\mathcal{M}^{-1}$	vector updates	inner products	solve with $\boldsymbol{E}$
Construction of C and $\boldsymbol{E}$	d	–	–	$d(d+1)/2$	–
Application of $\mathcal{P}$ or $\mathcal{P}^*$	–	–	d	d	1
Application of correction	–	–	d	d	1

needs exactly the same data, so no more storage is required. The construction of C needs d applications of the operator $\mathcal{A}$ but—as stated above—no application of the preconditioning operator $\mathcal{M}^{-1}$. Because $\boldsymbol{E}$ is Hermitian, $d(d+1)/2$ inner products have to be computed. One application of $\mathcal{P}$ or $\mathcal{P}^*$ requires d inner products, the solution of a linear system with the Hermitian d-by-d matrix $\boldsymbol{E}$, and d vector updates. We gather this information in Table 2.1.

Instead of correcting the last approximation $\widetilde{x}_n$, it is also possible to start with the corrected initial guess

$$x_0 = \mathcal{P}^*\widetilde{x}_0 + U\langle U, \mathcal{A}U\rangle_H^{-1}\langle U, b\rangle_H \tag{2.14}$$

and to use $\mathcal{P}^*$ as a right "preconditioner" (note that $\mathcal{P}^*$ is singular in general). The difference is mainly of algorithmic nature and will be described very briefly.

For an invertible linear operator $\mathcal{B} \in L(H)$, the right-preconditioned system $\mathcal{A}\mathcal{B}y = b$ can be solved for y and then the original solution can be obtained from $x = \mathcal{B}y$. Instead of x_0, the initial guess $y_0 := \mathcal{B}^{-1}x_0$ is used and the initial residual $r_0 = b - \mathcal{A}\mathcal{B}y_0 = b - \mathcal{A}x_0$ equals the residual of the unpreconditioned system. Then the iterates

$$y_n = y_0 + z_n \quad \text{with} \quad z_n \in \mathcal{K}_n(\mathcal{A}\mathcal{B}, r_0)$$

and $x_n := \mathcal{B}y_n = x_0 + \mathcal{B}z_n$ are constructed such that the residual $r_n = b - \mathcal{A}\mathcal{B}y_n = b - \mathcal{A}x_n$ is minimal in $\|\cdot\|_H$. If the operator $\mathcal{A}\mathcal{B}$ is self-adjoint, the MINRES method can again be used to solve this minimization problem. Note that y_0 is not needed and will never be computed explicitly. The right-preconditioning can of course be combined with a positive definite preconditioner as described in the introduction of Section 2.

We now take a closer look at the case $\mathcal{B} = \mathcal{P}^*$ which differs from the above description because $\mathcal{P}^*$ is not invertible in general. However, even if the right-preconditioned system is not consistent (i.e., $b \notin \operatorname{range}(\mathcal{A}\mathcal{P}^*)$), the above strategy can be used to solve the original linear system. With x_0 from equation (2.14), let us construct the iterates

$$x_n = x_0 + \mathcal{P}^*y_n \quad \text{with} \quad y_n \in \mathcal{K}_n(\mathcal{M}^{-1}\mathcal{A}\mathcal{P}^*, r_0) \tag{2.15}$$

such that the residual

$$r_n = \mathcal{M}^{-1}b - \mathcal{M}^{-1}\mathcal{A}x_n \tag{2.16}$$

ETNA
Kent State University
http://etna.math.kent.edu

has minimal $\|\cdot\|_{\mathcal{M}}$-norm. Inserting (2.15) and the definition of x_0 into equation (2.16) yields $r_n = \mathcal{M}^{-1}\mathcal{P}b - \mathcal{M}^{-1}\mathcal{P}\mathcal{A}y_n$ with $y_n \in \mathcal{K}_n(\mathcal{M}^{-1}\mathcal{A}\mathcal{P}^*, r_0) = \mathcal{K}_n(\mathcal{M}^{-1}\mathcal{P}\mathcal{A}, r_0)$. The minimization problem is thus the same as in the case where MINRES is applied to the linear system (2.13), and because both the operators and initial vectors coincide, the same Lanczos relation holds. Consequently the MINRES method can be applied to the right-preconditioned system

$$\mathcal{M}^{-1}\mathcal{A}\mathcal{P}^* y = \mathcal{M}^{-1}b, \quad x = \mathcal{P}^* y$$

with the corrected initial guess x_0 from equation (2.14). The key issue here is that the initial guess is treated as in (2.15). A deflated and preconditioned MINRES implementation following these ideas only needs the operator $\mathcal{P}^*$ and the corrected initial guess x_0. A correction step at the end is then unnecessary.

2.3. Ritz vector computation. So far we considered a single linear system and assumed that a basis for the construction of the projection used in the deflated system is given (e.g., some eigenvectors are given). We now turn to a sequence of preconditioned linear systems

$$\mathcal{M}_{(k)}^{-1}\mathcal{A}_{(k)}x^{(k)} = \mathcal{M}_{(k)}^{-1}b^{(k)}, \tag{2.17}$$

where $\mathcal{M}_{(k)}, \mathcal{A}_{(k)} \in L(H)$ are invertible and self-adjoint with respect to $\langle\cdot,\cdot\rangle_H$, $\mathcal{M}_{(k)}$ is positive definite, and $x^{(k)}, b^{(k)} \in H$ for $k \in \{1, \ldots, M\}$. To improve the readability we use subscript indices for operators and superscript indices for elements or tuples of the Hilbert space H. Such a sequence may arise from a time dependent problem or a nonlinear equation where solutions are approximated using Newton's method (cf. Section 3). We now assume that the operator $\mathcal{M}_{(k+1)}^{-1}\mathcal{A}_{(k+1)}$ differs only slightly from the previous operator $\mathcal{M}_{(k)}^{-1}\mathcal{A}_{(k)}$. Then it may be worthwhile to extract some eigenvector approximations from the Krylov subspace *and* the deflation subspace used in the solution of the kth system in order to accelerate convergence of the next system by deflating these extracted approximate eigenvectors.

For explaining the strategy in more detail, we omit the sequence index for a moment and always refer to the kth linear system if not specified otherwise. Assume that we used a tuple $U \in H^d$ whose elements form a $\langle\cdot,\cdot\rangle_{\mathcal{M}}$-orthonormal basis to set up the projection $\mathcal{P}_{\mathcal{M}}$ (cf. Definition 2.3) for the kth linear system (2.17). We then assume that the deflated and preconditioned MINRES method, with inner product $\langle\cdot,\cdot\rangle_{\mathcal{M}}$ and initial guess $\widetilde{x}_0$, has computed a satisfactory approximate solution after n steps. The MINRES method then constructs a basis of the Krylov subspace $\mathcal{K}_n(\mathcal{P}_{\mathcal{M}}\mathcal{M}^{-1}\mathcal{A}, r_0)$ where the initial residual is $r_0 = \mathcal{P}_{\mathcal{M}}\mathcal{M}^{-1}(b - \mathcal{A}\widetilde{x}_0)$. Due to the definition of the projection we know that $\mathcal{K}_n(\mathcal{P}_{\mathcal{M}}\mathcal{M}^{-1}\mathcal{A}, r_0) \perp_{\mathcal{M}} \operatorname{range}(U)$, and we now wish to compute approximate eigenvectors of $\mathcal{M}^{-1}\mathcal{A}$ in the subspace $S := \mathcal{K}_n(\mathcal{P}_{\mathcal{M}}\mathcal{M}^{-1}\mathcal{A}, r_0) \oplus \operatorname{range}(U)$. We can then pick some approximate eigenvectors according to the corresponding approximate eigenvalues and the approximation quality in order to construct a projection for the deflation of the $(k+1)$st linear system.

Let us recall the definition of Ritz pairs [36]:

DEFINITION 2.6. *Let $S \subseteq H$ be a finite dimensional subspace and let $\mathcal{B} \in L(H)$ be a linear operator. $(w, \mu) \in S \times \mathbb{C}$ is called a Ritz pair of $\mathcal{B}$ with respect to S and the inner product $\langle\cdot,\cdot\rangle$ if*

$$\mathcal{B}w - \mu w \perp_{\langle\cdot,\cdot\rangle} S.$$

The following lemma gives insight into how the Ritz pairs of the operator $\mathcal{M}^{-1}\mathcal{A}$ with respect to the Krylov subspace $\mathcal{K}_n(\mathcal{P}_{\mathcal{M}}\mathcal{M}^{-1}\mathcal{A}, r_0)$ and the deflation subspace $\operatorname{range}(U)$

ETNA
Kent State University
http://etna.math.kent.edu

can be obtained from data that are available when the MINRES method found a satisfactory approximate solution of the last linear system.

LEMMA 2.7. *Let the following assumptions hold:*

- *Let $\mathcal{M}, \mathcal{A}, U, \mathcal{P}_\mathcal{M}$ be defined as in Definition 2.3 and let $\langle U, U\rangle_\mathcal{M} = \boldsymbol{I}_d$.*
- *The Lanczos algorithm with inner product $\langle\cdot,\cdot\rangle_\mathcal{M}$ applied to the operator $\mathcal{P}_\mathcal{M}\mathcal{M}^{-1}\mathcal{A}$ and an initial vector $v \in \operatorname{range}(U)^{\perp_\mathcal{M}}$ proceeds to the nth iteration. The Lanczos relation is*

$$\mathcal{P}_\mathcal{M}\mathcal{M}^{-1}\mathcal{A}V_n = V_{n+1}\underline{\boldsymbol{T}}_n \tag{2.18}$$

with

$$\begin{aligned} V_{n+1} &= [v_1, \dots, v_{n+1}] \in H^{n+1}, \\ \langle V_{n+1}, V_{n+1}\rangle_\mathcal{M} &= \boldsymbol{I}_{n+1}, \qquad \text{and} \\ \underline{\boldsymbol{T}}_n &= \begin{bmatrix} \boldsymbol{T}_n \\ 0 \dots 0\ s_n \end{bmatrix} \in \mathbb{R}^{n+1,n}, \end{aligned}$$

where $s_n \in \mathbb{R}$ is positive and $\boldsymbol{T}_n \in \mathbb{R}^{n,n}$ is tridiagonal, symmetric, and real-valued.

- *Let $S := \mathcal{K}_n(\mathcal{P}_\mathcal{M}\mathcal{M}^{-1}\mathcal{A}, v) \oplus \operatorname{range}(U)$ and $w := [V_n, U]\widetilde{w} \in S$ for a $\widetilde{w} \in \mathbb{K}^{n+d}$.*

Then $(w, \mu) \in S \times \mathbb{R}$ is a Ritz pair of $\mathcal{M}^{-1}\mathcal{A}$ with respect to S and the inner product $\langle\cdot,\cdot\rangle_\mathcal{M}$ if and only if

$$\begin{bmatrix} \boldsymbol{T}_n + \boldsymbol{B}\boldsymbol{E}^{-1}\boldsymbol{B}^\mathrm{H} & \boldsymbol{B} \\ \boldsymbol{B}^\mathrm{H} & \boldsymbol{E} \end{bmatrix} \widetilde{w} = \mu\widetilde{w},$$

where $\boldsymbol{B} := \langle V_n, \mathcal{A}U\rangle_H$ and $\boldsymbol{E} := \langle U, \mathcal{A}U\rangle_H$.

Furthermore, the squared $\|\cdot\|_\mathcal{M}$-norm of the Ritz residual $\mathcal{M}^{-1}\mathcal{A}w - \mu w$ is

$$\left\|\mathcal{M}^{-1}\mathcal{A}w - \mu w\right\|_\mathcal{M}^2 = (\boldsymbol{G}\widetilde{w})^\mathrm{H} \begin{bmatrix} \boldsymbol{I}_{n+1} & \underline{\boldsymbol{B}} & 0 \\ \underline{\boldsymbol{B}}^\mathrm{H} & \boldsymbol{F} & \boldsymbol{E} \\ 0 & \boldsymbol{E} & \boldsymbol{I}_d \end{bmatrix} \boldsymbol{G}\widetilde{w},$$

where

$$\begin{aligned} \underline{\boldsymbol{B}} &= \langle V_{n+1}, \mathcal{A}U\rangle_H = \begin{bmatrix} \boldsymbol{B} \\ \langle v_{n+1}, \mathcal{A}U\rangle_H \end{bmatrix}, \\ \boldsymbol{F} &= \left\langle \mathcal{A}U, \mathcal{M}^{-1}\mathcal{A}U\right\rangle_H, \qquad \text{and} \\ \boldsymbol{G} &= \begin{bmatrix} \underline{\boldsymbol{T}}_n - \mu\underline{\boldsymbol{I}}_n & 0 \\ \boldsymbol{E}^{-1}\boldsymbol{B}^\mathrm{H} & \boldsymbol{I}_d \\ 0 & -\mu\boldsymbol{I}_d \end{bmatrix} \quad \text{with} \quad \underline{\boldsymbol{I}}_n = \begin{bmatrix} \boldsymbol{I}_n \\ 0 \end{bmatrix}. \end{aligned}$$

Proof. (w, μ) is a Ritz pair of $\mathcal{M}^{-1}\mathcal{A}$ with respect to $S = \operatorname{range}([V_n, U])$ and the inner product $\langle\cdot,\cdot\rangle_\mathcal{M}$ if and only if

$$\begin{aligned} & \mathcal{M}^{-1}\mathcal{A}w - \mu w \perp_\mathcal{M} S \\ \iff\quad & \left\langle s, \mathcal{M}^{-1}\mathcal{A}w - \mu w\right\rangle_\mathcal{M} = 0 \quad \forall s \in S \\ \iff\quad & \left\langle [V_n, U], (\mathcal{M}^{-1}\mathcal{A} - \mu\mathcal{I})[V_n, U]\right\rangle_\mathcal{M}\widetilde{w} = 0 \\ \iff\quad & \left\langle [V_n, U], \mathcal{M}^{-1}\mathcal{A}[V_n, U]\right\rangle_\mathcal{M}\widetilde{w} = \mu\langle [V_n, U], [V_n, U]\rangle_\mathcal{M}\widetilde{w} \\ \iff\quad & \left\langle [V_n, U], \mathcal{M}^{-1}\mathcal{A}[V_n, U]\right\rangle_\mathcal{M}\widetilde{w} = \mu\widetilde{w}, \end{aligned}$$

ETNA
Kent State University
http://etna.math.kent.edu

where the last equivalence follows from the orthonormality of U and V_n and the fact that $\operatorname{range}(U) \perp_{\mathcal{M}} \mathcal{K}_n(\mathcal{P}_{\mathcal{M}}\mathcal{M}^{-1}\mathcal{A}, v) = \operatorname{range}(V_n)$. We decompose the left-hand side as

$$\left\langle [V_n, U], \mathcal{M}^{-1}\mathcal{A}[V_n, U] \right\rangle_{\mathcal{M}} = \begin{bmatrix} \left\langle V_n, \mathcal{M}^{-1}\mathcal{A}V_n \right\rangle_{\mathcal{M}} & \left\langle V_n, \mathcal{M}^{-1}\mathcal{A}U \right\rangle_{\mathcal{M}} \\ \left\langle U, \mathcal{M}^{-1}\mathcal{A}V_n \right\rangle_{\mathcal{M}} & \left\langle U, \mathcal{M}^{-1}\mathcal{A}U \right\rangle_{\mathcal{M}} \end{bmatrix}.$$

The Lanczos relation (2.18) is equivalent to

$$\mathcal{M}^{-1}\mathcal{A}V_n = V_{n+1}\underline{\boldsymbol{T}}_n + \mathcal{M}^{-1}\mathcal{A}U\langle U, \mathcal{A}U\rangle_H^{-1}\langle \mathcal{A}U, V_n\rangle_H, \tag{2.19}$$

from which we can conclude with the $\langle\cdot,\cdot\rangle_{\mathcal{M}}$-orthonormality of $[V_{n+1}, U]$ that

$$\begin{aligned} \left\langle V_n, \mathcal{M}^{-1}\mathcal{A}V_n \right\rangle_{\mathcal{M}} &= \langle V_n, V_{n+1}\rangle_{\mathcal{M}}\underline{\boldsymbol{T}}_n + \left\langle V_n, \mathcal{M}^{-1}\mathcal{A}U \right\rangle_{\mathcal{M}}\langle U, \mathcal{A}U\rangle_H^{-1}\langle \mathcal{A}U, V_n\rangle_H \\ &= \boldsymbol{T}_n + \langle V_n, \mathcal{A}U\rangle_H\langle U, \mathcal{A}U\rangle_H^{-1}\langle \mathcal{A}U, V_n\rangle_H. \end{aligned}$$

The characterization of Ritz pairs is complete by recognizing that

$$\begin{aligned} &\boldsymbol{B} = \left\langle V_n, \mathcal{M}^{-1}\mathcal{A}U \right\rangle_{\mathcal{M}} = \langle V_n, \mathcal{A}U\rangle_H = \left\langle U, \mathcal{M}^{-1}\mathcal{A}V_n \right\rangle_{\mathcal{M}}^{\mathrm{H}} \qquad \text{and} \\ &\boldsymbol{E} = \left\langle U, \mathcal{M}^{-1}\mathcal{A}U \right\rangle_{\mathcal{M}} = \langle U, \mathcal{A}U\rangle_H. \end{aligned}$$

Only the equation for the residual norm remains to be shown. Therefore we compute with (2.19)

$$\begin{aligned} \mathcal{M}^{-1}\mathcal{A}w - \mu w &= \mathcal{M}^{-1}\mathcal{A}[V_n, U]\widetilde{w} - \mu[V_n, U]\widetilde{w} \\ &= [V_{n+1}, \mathcal{M}^{-1}\mathcal{A}U, U]\begin{bmatrix} \underline{\boldsymbol{T}}_n - \mu\underline{\boldsymbol{I}}_n & 0 \\ \boldsymbol{E}^{-1}\boldsymbol{B}^{\mathrm{H}} & \boldsymbol{I}_d \\ 0 & -\mu\boldsymbol{I}_d \end{bmatrix}\widetilde{w} \\ &= [V_{n+1}, \mathcal{M}^{-1}\mathcal{A}U, U]\boldsymbol{G}\widetilde{w}. \end{aligned}$$

The squared residual $\|\cdot\|_{\mathcal{M}}$-norm thus is

$$\left\|\mathcal{M}^{-1}\mathcal{A}w - \mu w\right\|_{\mathcal{M}}^2 = (\boldsymbol{G}\widetilde{w})^{\mathrm{H}}\left\langle [V_{n+1}, \mathcal{M}^{-1}\mathcal{A}U, U], [V_{n+1}, \mathcal{M}^{-1}\mathcal{A}U, U] \right\rangle_{\mathcal{M}}\boldsymbol{G}\widetilde{w},$$

where $\left\langle [V_{n+1}, \mathcal{M}^{-1}\mathcal{A}U, U], [V_{n+1}, \mathcal{M}^{-1}\mathcal{A}U, U] \right\rangle_{\mathcal{M}} = \begin{bmatrix} \boldsymbol{I}_{n+1} & \underline{\boldsymbol{B}} & 0 \\ \underline{\boldsymbol{B}}^{\mathrm{H}} & \boldsymbol{F} & \boldsymbol{E} \\ 0 & \boldsymbol{E} & \boldsymbol{I}_d \end{bmatrix}$ can be shown with the same techniques as above. ∎

REMARK 2.8. Lemma 2.7 also holds for the (rare) case that $\mathcal{K}_n(\mathcal{P}_{\mathcal{M}}\mathcal{M}^{-1}\mathcal{A}, v)$ is an invariant subspace of $\mathcal{P}_{\mathcal{M}}\mathcal{M}^{-1}\mathcal{A}$ which we excluded for readability reasons. The Lanczos relation (2.18) in this case is $\mathcal{P}_{\mathcal{M}}\mathcal{M}^{-1}\mathcal{A}V_n = V_n\boldsymbol{T}_n$, which does not change the result.

REMARK 2.9. Instead of using Ritz vectors for deflation, alternative approximations to eigenvectors are possible. An obvious choice are harmonic Ritz pairs $(w, \mu) \in S \times \mathbb{C}$ such that

$$\mathcal{B}w - \mu w \perp_{\langle\cdot,\cdot\rangle} \mathcal{B}S\,; \tag{2.20}$$

see [31, 36, 52]. However, in numerical experiments no significant difference between regular and harmonic Ritz pairs could be observed; see Remark 3.6 in Section 3.3.

Lemma 2.7 shows how a Lanczos relation for the operator $\mathcal{P}_{\mathcal{M}}\mathcal{M}^{-1}\mathcal{A}$ (that can be generated implicitly in the deflated and preconditioned MINRES algorithm, cf. the end of Section 2.2) can be used to obtain Ritz pairs of the "undeflated" operator $\mathcal{M}^{-1}\mathcal{A}$. An algorithm for the solution of the sequence of linear systems (2.17) as described in the beginning of this section is given in Algorithm 2.1. In addition to the Ritz vectors, this algorithm can include auxiliary deflation vectors $Y^{(k)}$.

ETNA
Kent State University
http://etna.math.kent.edu

Algorithm 2.1 Algorithm for the solution of the sequence of linear systems (2.17).

Input: For $k \in \{1, \dots, M\}$ we have:
- $\mathcal{M}_{(k)} \in L(H)$ is $\langle \cdot, \cdot \rangle_H$-self-adjoint and positive-definite. ▷ preconditioner
- $\mathcal{A}_{(k)} \in L(H)$ is $\langle \cdot, \cdot \rangle_H$-self-adjoint. ▷ operator
- $b^{(k)}, x_0^{(k)} \in H$. ▷ right hand side and initial guess
- $Y^{(k)} \in H^{l_{(k)}}$ for $l_{(k)} \in \mathbb{N}_0$. ▷ auxiliary deflation vectors (may be empty)

1: $W = [\,] \in H^0$ ▷ no Ritz vectors available in first step
2: **for** $k = 1 \to M$ **do**
3: $U =$ orthonormal basis of $\operatorname{span}[W, Y^{(k)}]$ with respect to $\langle \cdot, \cdot \rangle_{\mathcal{M}_{(k)}}$.
4: $C = \mathcal{A}_{(k)} U$, $\boldsymbol{E} = \langle U, C \rangle_H$ ▷ $\mathcal{P}^*$ as in Lemma 2.4
5: $x_0 = \mathcal{P}^* x_0^{(k)} + U \boldsymbol{E}^{-1} \langle U, b^{(k)} \rangle_H$ ▷ corrected initial guess
6: $x_n^{(k)}, V_{n+1}, \underline{\boldsymbol{T}}_n, \boldsymbol{B} = \text{MINRES}(\mathcal{A}_{(k)}, b_{(k)}, \mathcal{M}_{(k)}^{-1}, \mathcal{P}^*, x_0, \varepsilon)$
MINRES is applied to $\mathcal{M}_{(k)}^{-1} \mathcal{A}_{(k)} x^{(k)} = \mathcal{M}_{(k)}^{-1} b^{(k)}$ with inner product $\langle \cdot, \cdot \rangle_{\mathcal{M}_{(k)}}$, right preconditioner $\mathcal{P}^*$, initial guess x_0 and tolerance $\varepsilon > 0$; cf. Section 2.2. Then:
- The approximation $x_n^{(k)}$ fulfills $\left\| \mathcal{M}_{(k)}^{-1} b^{(k)} - \mathcal{M}_{(k)}^{-1} \mathcal{A}_{(k)} x_n^{(k)} \right\|_{\mathcal{M}_{(k)}} \leq \varepsilon$.
- The Lanczos relation $\mathcal{M}_{(k)}^{-1} \mathcal{A}_{(k)} \mathcal{P}^* V_n = V_{n+1} \underline{\boldsymbol{T}}_n$ holds.
- $\boldsymbol{B} = \langle V_n, C \rangle_H$ is generated as a byproduct of the application of $\mathcal{P}^*$.

7: $w_1, \dots, w_m, \mu_1, \dots, \mu_m, \rho_1, \dots, \rho_m = \text{Ritz}(U, V_{n+1}, \underline{\boldsymbol{T}}_n, \boldsymbol{B}, C, \boldsymbol{E}, \mathcal{M}_{(k)}^{-1})$
Ritz(...) computes the Ritz pairs (w_j, μ_j) for $j \in \{1, \dots, m\}$ of $\mathcal{M}_{(k)}^{-1} \mathcal{A}_{(k)}$ with respect to $\operatorname{span}[U, V_n]$ and the inner product $\langle \cdot, \cdot \rangle_{\mathcal{M}_{(k)}}$, cf. Lemma 2.7. Then:
- $w_1, \dots, w_m$ form a $\langle \cdot, \cdot \rangle_{\mathcal{M}_{(k)}}$-orthonormal basis of $\operatorname{span}[U, V_n]$.
- The residual norms $\rho_j = \left\| \mathcal{M}_{(k)}^{-1} \mathcal{A}_{(k)} w_j - \mu_j w_j \right\|_{\mathcal{M}_{(k)}}$ are also returned.

8: $W = [w_{i_1}, \dots, w_{i_d}]$ for pairwise distinct $i_1, \dots, i_d \in \{1, \dots, m\}$.
Pick d Ritz vectors according to Ritz value and residual norm.
9: **end for**

2.3.1. Selection of Ritz vectors. In step 8 of Algorithm 2.1, up to m Ritz vectors can be chosen for deflation in the next linear system. It is unclear which choice leads to optimal convergence. The convergence of MINRES is determined by the spectrum of the operator and the initial residual in an intricate way. In most applications one can only use rough convergence bounds of the type (2.6) which form the basis for certain heuristics. Popular choices include Ritz vectors corresponding to smallest- or largest-magnitude Ritz values or smallest Ritz residual norms. No general recipe can be expected.

2.3.2. Notes on the implementation. We now comment on the implementational side of the determination and utilization of Ritz pairs while solving a sequence of linear systems (cf. Algorithm 2.1). The solution of a single linear system with the deflated and preconditioned MINRES method was discussed in Section 2.2. Although the MINRES method is based on short recurrences due to the underlying Lanczos algorithm—and thus only needs storage for a few vectors—we still have to store the full Lanczos basis V_{n+1} for the determination of Ritz vectors and the Lanczos matrix $\underline{\boldsymbol{T}}_n \in \mathbb{R}^{n+1,n}$. The storage requirements of the tridiagonal Lanczos matrix are negligible while storing all Lanczos vectors may be costly. As customary for GMRES, this difficulty can be overcome by restarting the MINRES method after a fixed

ETNA
Kent State University
http://etna.math.kent.edu

number of iterations. This could be added trivially to Algorithm 2.1 as well by iterating lines 3 to 8 with a fixed maximum number of MINRES iterations for the same linear system and the last iterate as initial guess. In this case, the number n is interpreted not as the total number of MINRES iterations but as the number of MINRES iterations in a restart phase. As an alternative to restarting, Wang et al. [52] suggest to compute the Ritz vectors in cycles of fixed length s. At the end of each cycle, new Ritz vectors are computed from the previous Ritz vectors and the s Lanczos vectors from the current cycle. All but the last two Lanczos vectors are then dropped since they are not required for continuing the MINRES iteration. Therefore, the method in [52] is able to maintain global optimality of the approximate solution with respect to the entire Krylov subspace (in exact arithmetic), which may lead to faster convergence compared to restarted methods. Note that a revised RMINRES implementation with performance optimizations has been published in [28]. Both restarting and cycling thus provide a way to limit the memory requirements. However, due to the loss of information, the quality of computed Ritz vectors and thus their performance as recycling vectors typically deteriorates with both strategies. For example, this has been observed experimentally by Wang et al. [52], where recycling vectors from shorter cycles were less effective.

In the experiments in this manuscript, neither restarting nor cycling is necessary since the preconditioner sufficiently limits the number of iterations (cf. Section 3.3). Deflation can then be used to further improve convergence by directly addressing parts of the preconditioned operator's spectrum. An annotated version of the algorithm can be found in Algorithm 2.1. Note that the inner product matrix $\boldsymbol{B}$ is computed implicitly row-wise in each iteration of MINRES by the application of $\mathcal{P}^*$ to the last Lanczos vector v_n because this involves the computation of $\langle \mathcal{A}U, v_n \rangle = \boldsymbol{B}^{\mathrm{H}} e_n$.

2.3.3. Overall computational cost. An overview of the computational cost of one iteration of Algorithm 2.1 is given in Table 2.2. The computation of one iteration of Algorithm 2.1 with n MINRES steps and d deflation vectors involves $n + d + 1$ applications of the preconditioner $\mathcal{M}^{-1}$ and the operator $\mathcal{A}$. These steps are typically very costly and dominate the overall computation time. This is true for all variants of recycling Krylov subspace methods. With this in mind, we would like to take a closer look at the cost induced by the other elements of the algorithm. If the inner products are assumed Euclidean, their computation accounts for a total of $2N \times (d^2 + nd + 3d + 2n)$ FLOPs. If the selection strategy of Ritz vectors for recycling requires knowledge of the respective Ritz residuals, an additional $2N \times d^2$ FLOPs must be invested. The vector updates require $2N \times (3/2d^2 + 2nd + 5/2d + 7n)$ FLOPs, so in total, without computation of Ritz residuals, $2N \times (5/2d^2 + 3nd + 11/2d + 9n)$ FLOPs are required for one iteration of Algorithm 2.1 in addition to the operator applications.

Comparing the computational cost of the presented method with restarted or cycled methods is hardly possible. If the cycle length s in [52] equals the overall number of iterations n, then that method requires $2N \times (6d^2 + 3nd + 3d + 2)$ FLOPs for updating the recycling space. In practice, the methods show a different convergence behavior because $s \ll n$ and the involved projections differ; cf. Section 2.2.

Note that the orthonormalization in line 3 is redundant in exact arithmetic if only Ritz vectors are used and the preconditioner does not change. Further note that the orthogonalization requires the application of the operator $\mathcal{M}$, i.e., the inverse of the preconditioner. This operator is not known in certain cases, e.g., for the application of only a few cycles of an (algebraic) multigrid preconditioner. Orthogonalizing the columns of U with an inaccurate approximation of $\mathcal{M}$ (e.g., the original operator $\mathcal{B}$) will then make the columns of U formally orthonormal with respect to a different inner product. This may lead to wrong results in the Ritz value computation. A workaround in the popular case of (algebraic) multigrid preconditioners is to use so many cycles that $\mathcal{M} \approx \mathcal{B}$ is fulfilled with high accuracy. However, this typically leads

ETNA
Kent State University
http://etna.math.kent.edu

TABLE 2.2
Computational cost for one iteration of Algorithm 2.1 (lines 3–8) with n MINRES iterations and d deflation vectors. The number of computed Ritz vectors also is d. Operations that do not depend on the dimension $N := \dim H$ are neglected.

	Applications of $\mathcal{A}$	Applications of $\mathcal{M}^{-1}$	Applications of $\mathcal{M}$	Inner products	Vector updates
Orthogonalization	–	–	d	$d(d+1)/2$	$d(d+1)/2$
Setup of $\mathcal{P}^*$ and x_0	d	–	–	$d(d+3)/2$	d
n MINRES iterations	$n+1$	$n+1$	–	$n(d+2)+d$	$n(d+7)+d$
Comp. of Ritz vectors	–	d	–	–	$d(d+n)$
(Comp. of Ritz res. norms)	–	–	–	d^2	–

to a substantial increase in computational cost and, depending on the application, may defeat the original purpose of speeding up the Krylov convergence by recycling.

Similarly, round-off errors may lead to a loss of orthogonality in the Lanczos vectors and thus to inaccuracies in the computed Ritz pairs. Details on this are given in Remark 3.8.

3. Application to nonlinear Schrödinger problems. Given an open domain $\Omega \subseteq \mathbb{R}^{\{2,3\}}$, nonlinear Schrödinger operators are typically derived from the minimization of the Gibbs energy in a corresponding physical system and have the form

$$\begin{aligned} &\mathcal{S} : X \to Y, \\ &\mathcal{S}(\psi) := (\mathcal{K} + V + g|\psi|^2)\psi \quad \text{in } \Omega, \end{aligned} \tag{3.1}$$

with $X \subseteq L^2(\Omega)$ being the natural energy space of the problem and $Y \subseteq L^2(\Omega)$. If the domain is bounded, then the space X may incorporate boundary conditions appropriate to the physical setting. The linear operator $\mathcal{K}$ is assumed to be self-adjoint and positive-semidefinite with respect to $\langle \cdot, \cdot \rangle_{L^2(\Omega)}$, $V : \Omega \to \mathbb{R}$ is a given scalar potential, and $g > 0$ is a given nonlinearity parameter. A state $\hat{\psi} : \Omega \to \mathbb{C}$ is called a solution of the nonlinear Schrödinger equation if

$$\mathcal{S}(\hat{\psi}) = 0. \tag{3.2}$$

Generally, one is only interested in nontrivial solutions $\hat{\psi} \not\equiv 0$. The function $\hat{\psi}$ is often referred to as *order parameter* and its magnitude $|\hat{\psi}|^2$ typically describes a particle density or, more generally, a probability distribution. Note that, because of

$$\mathcal{S}(\exp\{\mathrm{i}\chi\}\psi) = \exp\{\mathrm{i}\chi\}\mathcal{S}(\psi), \tag{3.3}$$

one solution $\hat{\psi} \in X$ is really just a representative of the physically equivalent solutions $\{\exp\{\mathrm{i}\chi\}\hat{\psi} : \chi \in \mathbb{R}\}$.

For the numerical solution of (3.2), Newton's method is popular for its fast convergence in a neighborhood of a solution: given a good-enough initial guess ψ_0, the Newton process generates a sequence of iterates ψ_k which converges superlinearly towards a solution $\hat{\psi}$ of (3.2). In each step k of Newton's method, a linear system with the Jacobian

$$\begin{aligned} &\mathcal{J}(\psi) : X \to Y, \\ &\mathcal{J}(\psi)\phi := \left(\mathcal{K} + V + 2g|\psi|^2\right)\phi + g\psi^2\overline{\phi}. \end{aligned} \tag{3.4}$$

of $\mathcal{S}$ at ψ_k needs to be solved. Despite the fact that states ψ are generally complex-valued, $\mathcal{J}(\psi)$ is linear only if X and Y are defined as vector spaces over the field $\mathbb{R}$ with the corresponding

ETNA
Kent State University
http://etna.math.kent.edu

inner product

$$\langle \cdot, \cdot \rangle_{\mathbb{R}} := \Re \langle \cdot, \cdot \rangle_{L^2(\Omega)} . \tag{3.5}$$

This matches the notion that the specific complex argument of the order parameter is of no physical relevancy since $|r \exp\{\mathrm{i}\alpha\}\psi|^2 = |r\psi|^2$ for all $r, \alpha \in \mathbb{R}$, $\psi \in X$ (compare with (3.3)).

Moreover, the work in [42] gives a representation of adjoints of operators of the form (3.4), from which one can derive the following assertion:

COROLLARY 3.1. *For any given $\psi \in Y$, the Jacobian operator $\mathcal{J}(\psi)$ (3.4) is self-adjoint with respect to the inner product* (3.5).

An important consequence of the independence of states of the complex argument (3.3) is the fact that solutions of equation (3.1) form a smooth manifold in X. Therefore, the linearization (3.4) in solutions always has a nontrivial kernel. Indeed, for any $\psi \in X$

$$\mathcal{J}(\psi)(\mathrm{i}\psi) = \left(\mathcal{K} + V + 2g|\psi|^2\right)(\mathrm{i}\psi) - g\mathrm{i}\psi^2\overline{\psi} = \mathrm{i}\left(\mathcal{K} + V + g|\psi|^2\right)\psi = \mathrm{i}\mathcal{S}(\psi), \tag{3.6}$$

so for nontrivial solutions $\hat{\psi} \in X$, $\psi \not\equiv 0$, $\mathcal{S}(\hat{\psi}) = 0$, the dimensionality of the kernel of $\mathcal{J}(\psi)$ is at least 1.

Besides the fact that there is always a zero eigenvalue in a solution $\hat{\psi}$ and that all eigenvalues are real, not much more can be said about the spectrum; in general, $\mathcal{J}(\psi)$ is indefinite. The definiteness depends entirely on the state ψ. For ψ being a solution to (3.1), it is said to be stable or unstable depending whether or not $\mathcal{J}(\psi)$ has negative eigenvalues. Typically, solutions with low Gibbs energies tend to be stable whereas highly energetic solutions tend to be unstable. For physical systems in practice, it is uncommon to see more than ten negative eigenvalues for a given solution state.

3.1. Principal problems for the numerical solution. While the numerical solution of nonlinear systems itself is challenging, the presence of a singularity in a solution as in (3.6) adds two major obstacles for using Newton's method.

- Newton's method is guaranteed to converge towards a solution $\hat{\psi}$ Q-superlinearly in the area of attraction only if $\hat{\psi}$ is nondegenerate, i.e., the Jacobian in $\hat{\psi}$ is regular. If the Jacobian operator does have a singularity, then only linear convergence can be guaranteed.
- While no linear system has to be solved with the exactly singular $\mathcal{J}(\hat{\psi})$, the Jacobian operator close to the solution $\mathcal{J}(\hat{\psi} + \delta\psi)$ will have at least one eigenvalue of small magnitude, i.e., the Jacobian system becomes ill-conditioned when approaching a solution.

Several approaches have been suggested to deal with this situation; for a concise survey of the matter, see [21]. One of the most used strategies is *bordering*, which suggests extending the original problem $\mathcal{S}(\psi) = 0$ by a so-called *phase condition* to pin down the redundancy [1],

$$0 = \tilde{\mathcal{S}}(\psi, \lambda) := \begin{bmatrix} \mathcal{S}(\psi) + \lambda y \\ p(x) \end{bmatrix} . \tag{3.7}$$

If y and $p(\cdot)$ are chosen according to some well-understood criteria [23], the Jacobian systems can be shown to be well-conditioned throughout the Newton process. Moreover, the bordering can be chosen in such a way that the linearization of the extended system is self-adjoint in the extended scalar product if the linearization of the original problem is also self-adjoint. This method has been applied to the specialization of the Ginzburg–Landau equations (3.10) before [42] and naturally generalizes to nonlinear Schrödinger equations in the same way.

ETNA
Kent State University
http://etna.math.kent.edu

One major disadvantage of the bordering approach, however, is that it is not clear how to precondition the extended system even if a good preconditioner for the original problem is known.

In the particular case of nonlinear Schrödinger equations, the loss of speed of convergence is less severe than in more general settings. Note that there would be no slowdown at all if the Newton update $\delta\psi$, given by

$$\mathcal{J}(\psi)\delta\psi = -\mathcal{S}(\psi), \tag{3.8}$$

was consistently orthogonal to the null space $\mathrm{i}\hat{\psi}$ close to a solution $\hat{\psi}$. While this is not generally true, one is at least in the situation that the Newton update can never be an exact multiple of the direction of the approximate null space $\mathrm{i}\psi$. This is because

$$\mathcal{J}(\psi)(\alpha\mathrm{i}\psi) = -\mathcal{S}(\psi), \quad \alpha \in \mathbb{R},$$

together with (3.6), is equivalent to

$$\alpha\mathrm{i}\mathcal{S}(\psi) = -\mathcal{S}(\psi),$$

which can only be fulfilled if $\mathcal{S}(\psi) = 0$, i.e., if ψ is already a solution.

Consequently, loss of Q-superlinear convergence is hardly ever observed in numerical experiments. Figure 3.1, for example, shows the Newton residual for the two- and three-dimensional test setups, both with the standard formulation and with the bordering (3.7) as proposed in [42]. Of course, the Newton iterates follow different trajectories, but the important thing to note is that in both plain and bordered formulation, the speed of convergence close to the solution is comparable.

The more severe restriction lies in the numerical difficulty of solving the Jacobian systems in each Newton step due to the increasing ill-posedness of the problem as described above. However, although the Jacobian has a nontrivial near-null space close to a solution, the problem is well-defined at all times. This is because, by self-adjointness, its left near-null space coincides with the right near-null space, $\operatorname{span}\{\mathrm{i}\hat{\psi}\}$, and the right-hand-side in (3.8), $-\mathcal{S}(\psi)$, is orthogonal to $\mathrm{i}\psi$ for any ψ:

$$\begin{aligned}\langle \mathrm{i}\psi, S(\psi)\rangle_{\mathbb{R}} &= \langle \mathrm{i}\psi, \mathcal{K}(\psi)\rangle_{\mathbb{R}} + \langle \mathrm{i}\psi, V(\psi)\rangle_{\mathbb{R}} + \left\langle \mathrm{i}\psi, g|\psi|^2\psi\right\rangle_{\mathbb{R}} \\ &= \Re\left(\mathrm{i}\langle\psi, \mathcal{K}\psi\rangle_2\right) + \Re\left(\mathrm{i}\langle\psi, V\psi\rangle_2\right) + \Re\left(g\mathrm{i}\left\langle|\psi|^2, |\psi|^2\right\rangle_2\right) = 0.\end{aligned} \tag{3.9}$$

The numerical problem is hence caused only by the fact that one eigenvalue approaches the origin as the Newton iterates approach a solution. The authors propose to handle this difficulty at the level of the linear solves for the Newton updates using the deflation framework developed in Section 2.

3.2. The Ginzburg–Landau equation. One important instance of nonlinear Schrödinger equations (3.1) is the Ginzburg–Landau equation that models supercurrent density for extreme-type-II superconductors. Given an open, bounded domain $\Omega \subseteq \mathbb{R}^{\{2,3\}}$, the equations are

$$0 = \begin{cases} \mathcal{K}\psi - \psi(1 - |\psi|^2) & \text{in } \Omega, \\ \mathbf{n}\cdot(-\mathrm{i}\boldsymbol{\nabla} - \boldsymbol{A})\psi & \text{on } \partial\Omega. \end{cases} \tag{3.10}$$

The operator $\mathcal{K}$ is defined as

$$\begin{aligned}&\mathcal{K}\colon X \to Y, \\ &\mathcal{K}\phi := (-\mathrm{i}\boldsymbol{\nabla} - \boldsymbol{A})^2\phi,\end{aligned}$$

ETNA
Kent State University
http://etna.math.kent.edu

with the magnetic vector potential $\boldsymbol{A} \in H^2_{\mathbb{R}^d}(\Omega)$ [7]. The operator $\mathcal{K}$ describes the energy of a charged particle under the influence of a magnetic field $\boldsymbol{B} = \boldsymbol{\nabla} \times \boldsymbol{A}$ and can be shown to be Hermitian and positive-semidefinite; the eigenvalue 0 is assumed only for $\boldsymbol{A} \equiv \boldsymbol{0}$ [43]. Solutions $\hat{\psi}$ of (3.10) describe the density $|\hat{\psi}|^2$ of electric charge carriers and fulfill the inequalities $0 \leq |\hat{\psi}|^2 \leq 1$ pointwise [7]. For two-dimensional domains, they typically exhibit isolated zeros referred to as *vortices*; in three dimensions, lines of zeros are the typical solution pattern; see Figure 3.2.

3.2.1. Discretization. For the numerical experiments in this paper, a finite-volume-type discretization is employed [6, 43]. Let $\Omega^{(h)}$ be a discretization of Ω with a triangulation $\{T_i\}_{i=1}^m$, $\bigcup_{i=1}^m T_i = \Omega^{(h)}$, and the node-centered Voronoi tessellation $\{\Omega_k\}_{k=1}^n$, $\bigcup_{k=1}^n \Omega_k = \Omega^{(h)}$. Let further $\boldsymbol{e}_{i,j}$ denote the edge between two nodes i, j. The discretized problem is then to find $\psi^{(h)} \in \mathbb{C}^n$ such that

$$\forall k \in \{1, \ldots, n\}: \quad 0 = \left(S^{(h)}\psi^{(h)}\right)_k := \left(K^{(h)}\psi^{(h)}\right)_k - \psi_k^{(h)}\left(1 - |\psi_k^{(h)}|^2\right),$$

where the discrete kinetic energy operator $K^{(h)}$ is defined by

$$\forall \phi^{(h)}, \psi^{(h)} \in \mathbb{C}^n: \quad \left\langle K^{(h)}\psi^{(h)}, \phi^{(h)} \right\rangle = \sum_{\text{edges } \boldsymbol{e}_{i,j}} \alpha_{i,j}\left[\left(\psi_i^{(h)} - U_{i,j}\psi_j^{(h)}\right)\overline{\phi}_i^{(h)} + \left(\psi_j^{(h)} - \overline{U_{i,j}}\psi_i^{(h)}\right)\overline{\phi}_j^{(h)}\right]$$

with the discrete inner product

$$\left\langle \psi^{(h)}, \phi^{(h)} \right\rangle := \sum_{k=1}^n |\Omega_k|\, \psi_k^{(h)}\overline{\phi}_k^{(h)}$$

and edge coefficients $\alpha_{i,j} \in \mathbb{R}$ [43]. The magnetic vector potential $\boldsymbol{A}$ is incorporated in the so-called *link variables*,

$$U_{i,j} := \exp\left(-\mathrm{i}\int_{\boldsymbol{x}_j}^{\boldsymbol{x}_i} \boldsymbol{e}_{i,j} \cdot \boldsymbol{A}(\boldsymbol{w})\, \mathrm{d}\boldsymbol{w}\right)$$

along the edges $\boldsymbol{e}_{i,j}$ of the triangulation.

REMARK 3.2. In matrix form, the operator $K^{(h)}$ is represented as the product $K^{(h)} = D^{-1}\widehat{K}$ of the diagonal matrix D^{-1}, $D_{i,i} = |\Omega_i|$ and a Hermitian matrix $\widehat{K}$.

This discretization preserves a number of invariants of the problem, e.g., gauge invariance of the type $\tilde{\psi} := \exp\{\mathrm{i}\chi\}\psi$, $\tilde{\boldsymbol{A}} := \boldsymbol{A} + \nabla\chi$ with a given $\chi \in C^1(\Omega)$. Moreover, the discretized energy operator $K^{(h)}$ is Hermitian and positive-definite [43]. Analogous to (3.4), the discretized Jacobian operator at $\psi^{(h)}$ is defined by

$$J^{(h)}(\psi^{(h)}) : \mathbb{C}^n \to \mathbb{C}^n,$$
$$J^{(h)}(\psi^{(h)})\phi^{(h)} := \left(K^{(h)} - 1 + 2|\psi^{(h)}|^2\right)\phi^{(h)} + (\psi^{(h)})^2\overline{\phi^{(h)}},$$

where the vector-vector products are interpreted entry-wise. The discrete Jacobian is self-adjoint with respect to the discrete inner product

$$\left\langle \psi^{(h)}, \phi^{(h)} \right\rangle_{\mathbb{R}} := \Re\left(\sum_{k=1}^n |\Omega_k|\, \overline{\psi}_k^{(h)}\phi_k^{(h)}\right), \tag{3.11}$$

and the statements (3.6), (3.9) about the null space carry over from the continuous formulation.

ETNA
Kent State University
http://etna.math.kent.edu

REMARK 3.3 (Real-valued formulation). There is a vector space isomorphism between $\mathbb{R}^{2n}$ and $\mathbb{C}^n$ as vector spaces over $\mathbb{R}$ given by the basis mapping $\alpha : \mathbb{C}^n \to \mathbb{R}^{2n}$,

$$\alpha(e_j^{(n)}) = e_j^{(2n)}, \quad \alpha(\mathrm{i}e_j^{(n)}) = e_{n+j}^{(2n)}.$$

In particular, note that the dimensionality of $\mathbb{C}^n_{\mathbb{R}}$ is $2n$. The isomorphism α is also isometric with the natural inner product $\langle \cdot, \cdot \rangle_{\mathbb{R}}$ of $\mathbb{C}^n_{\mathbb{R}}$ since for any given pair $\phi, \psi \in \mathbb{C}^n$ one has

$$\left\langle \begin{pmatrix} \Re\phi \\ \Im\phi \end{pmatrix}, \begin{pmatrix} \Re\psi \\ \Im\psi \end{pmatrix} \right\rangle = \langle \Re\phi, \Re\psi \rangle + \langle \Im\phi, \Im\psi \rangle = \langle \phi, \psi \rangle_{\mathbb{R}}.$$

Moreover, linear operators over $\mathbb{C}^n_{\mathbb{R}}$ generally have the form $L\psi = A\psi + B\overline{\psi}$ with some $A, B \in \mathbb{C}^{n\times n}$, and because of

$$Lw = \lambda w \quad \Leftrightarrow \quad (\alpha L \alpha^{-1})\alpha w = \lambda \alpha w,$$

the eigenvalues also exactly convey to its real-valued image $\alpha L \alpha^{-1}$.

This equivalence can be relevant in practice as quite commonly the original complex-valued problem in $\mathbb{C}^n$ is implemented in terms of $\mathbb{R}^{2n}$. Using the natural inner product in this space will yield the expected results without having to take particular care of the inner product.

3.3. Numerical experiments. The numerical experiments are performed with the following two setups.

TEST SETUP 1 (2D). The circle $\Omega_{\text{2D}} := \{x \in \mathbb{R}^2 : \|\boldsymbol{x}\|_2 < 5\}$ and the magnetic vector potential $\boldsymbol{A}(\boldsymbol{x}) := \boldsymbol{m} \times (\boldsymbol{x} - \boldsymbol{x}_0)/\|\boldsymbol{x} - \boldsymbol{x}_0\|^3$ with $\boldsymbol{m} := (0, 0, 1)^{\mathrm{T}}$ and $\boldsymbol{x}_0 := (0, 0, 5)^{\mathrm{T}}$, corresponding to the magnetic field generated by a dipole at $\boldsymbol{x}_0$ with orientation $\boldsymbol{m}$. A Delaunay triangulation for this domain with 3299 nodes was created using *Triangle* [44]. With the discrete equivalent of $\psi_0(\boldsymbol{x}) = \cos(\pi y)$ as initial guess, the Newton process converges after 27 iterations with a residual of less than 10^{-10} in the discretized norm; see Figure 3.1. The final state is illustrated in Figure 3.2.

TEST SETUP 2 (3D). The three-dimensional L-shape

$$\Omega_{\text{3D}} := \{\boldsymbol{x} \in \mathbb{R}^3 : \|\boldsymbol{x}\|_\infty < 5\} \backslash \mathbb{R}^3_+,$$

discretized using Gmsh [18] with 72166 points. The chosen magnetic vector field is constant $\boldsymbol{B}_{\text{3D}}(\boldsymbol{x}) := 3^{-1/2}(1, 1, 1)^{\mathrm{T}}$, represented by the vector potential $\boldsymbol{A}_{\text{3D}}(\boldsymbol{x}) := \frac{1}{2}\boldsymbol{B}_{\text{3D}} \times \boldsymbol{x}$. With the discrete equivalent of $\psi_0(\boldsymbol{x}) = 1$, the Newton process converges after 22 iterations with a residual of less than 10^{-10} in the discretized norm; see Figure 3.1. The final state is illustrated in Figure 3.2.

All experimental results presented in this section can be reproduced from the data published with the free and open source Python packages *KryPy* [15] and *PyNosh* [16]. *KryPy* contains an implementation of deflated Krylov subspace methods; e.g., Algorithm 2.1. *PyNosh* provides solvers for nonlinear Schrödinger equations including the above test cases.

For both setups, Newton's method was used and the linear systems (3.8) were solved using MINRES to exploit self-adjointness of $J^{(h)}$. Note that it is critical here to use the natural inner product of the system (3.11). All of the numerical experiments incorporate the preconditioner proposed in [43] that is shown to bound the number of Krylov iterations needed to reach a certain relative residual by a constant independent of the number n of unknowns in the system.

REMARK 3.4. Neither of the above test problems have initial guesses which sit in the cone of attraction of the solution they eventually converge to. As typical for local nonlinear solvers, the iterations which do not directly correspond with the final convergence are sensitive

ETNA
Kent State University
http://etna.math.kent.edu

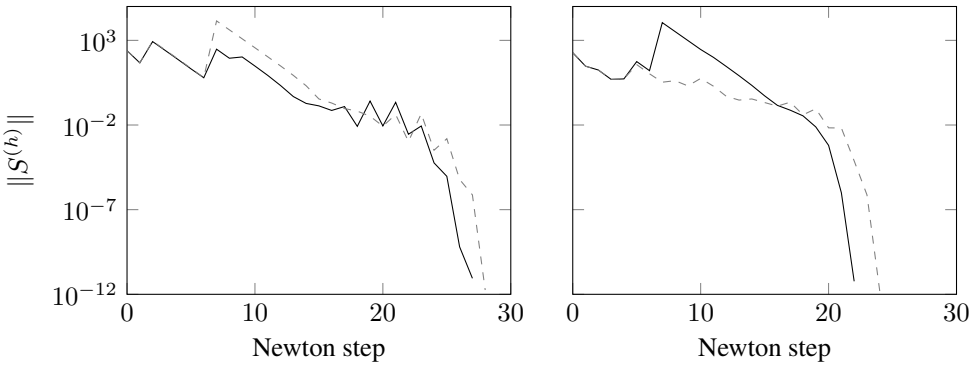
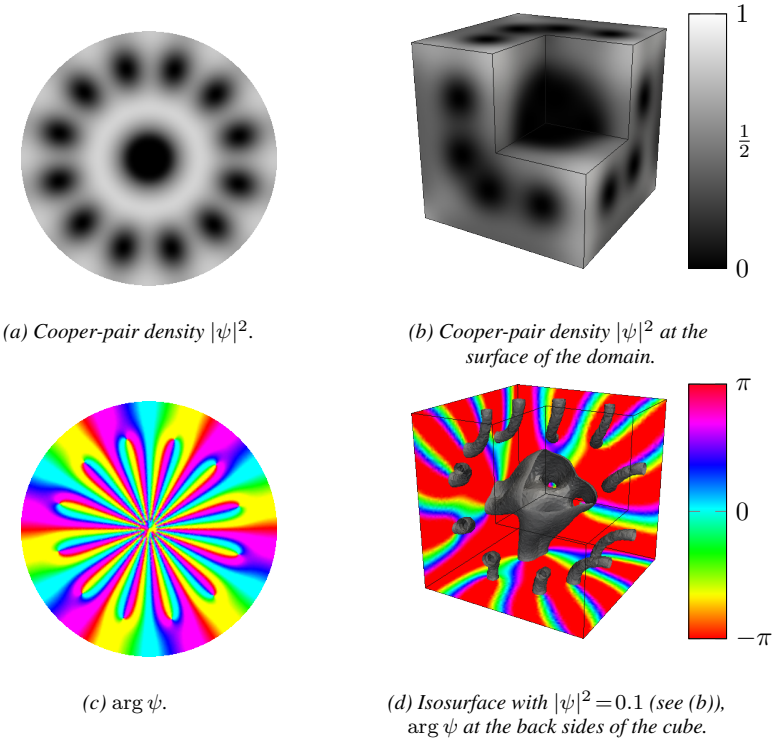


FIG. 3.1. *Newton residual history for the two-dimensional Test Setup 1 (left) and three-dimensional Test Setup 2 (right), each with bordering and without. With the initial guesses* $\psi_0^{2D}(\boldsymbol{x}) = \cos(\pi y)$ *and* $\psi_0^{3D}(\boldsymbol{x}) = 1$*, respectively, the Newton process delivered the solutions as highlighted in Figure 3.2 in 22 and 27 steps, respectively.*



(a) Cooper-pair density $|\psi|^2$.

(b) Cooper-pair density $|\psi|^2$ *at the surface of the domain.*

(c) $\arg\psi$.

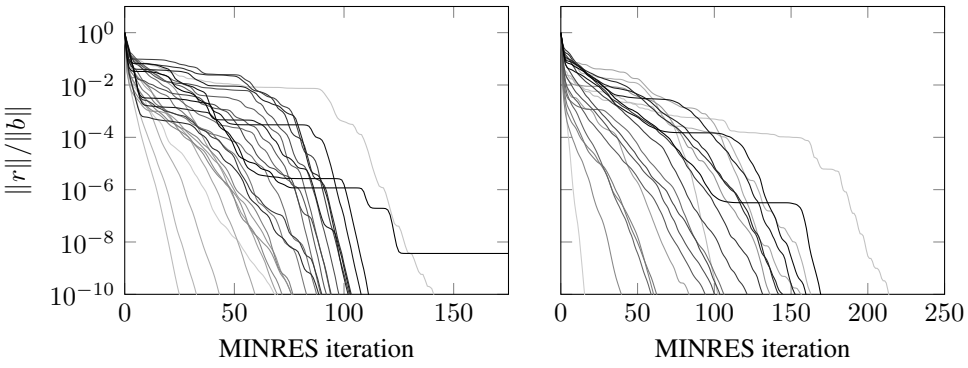
(d) Isosurface with $|\psi|^2 = 0.1$ *(see (b)),* $\arg\psi$ *at the back sides of the cube.*

FIG. 3.2. *Solutions of the test problems as found in the Newton process illustrated in Figure 3.1.*

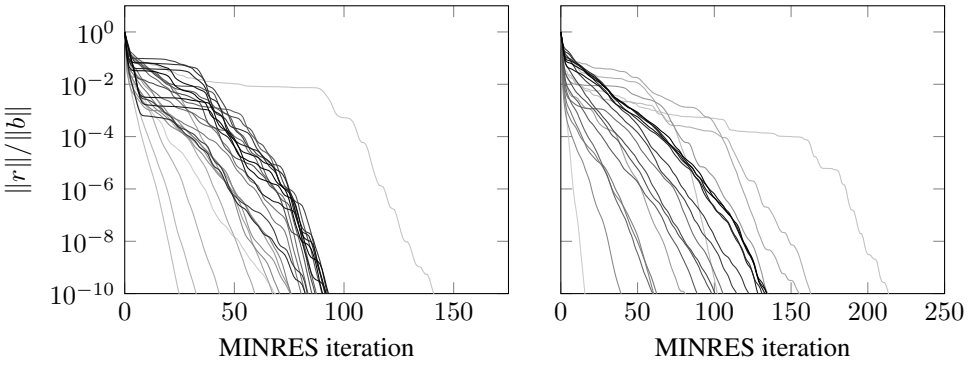
to effects introduced by the discretization or round-off errors. It will hence be difficult to reproduce precisely the shown solutions without exact information about the point coordinates in the discretization mesh. However, the same general convergence patterns were observed for numerous meshes and initial states; the presented solutions shall serve as examples thereof.

Figure 3.3 displays the relative residuals for all Newton steps in both the two- and the three-dimensional setup. Note that the residual curves late in the Newton process (dark gray)

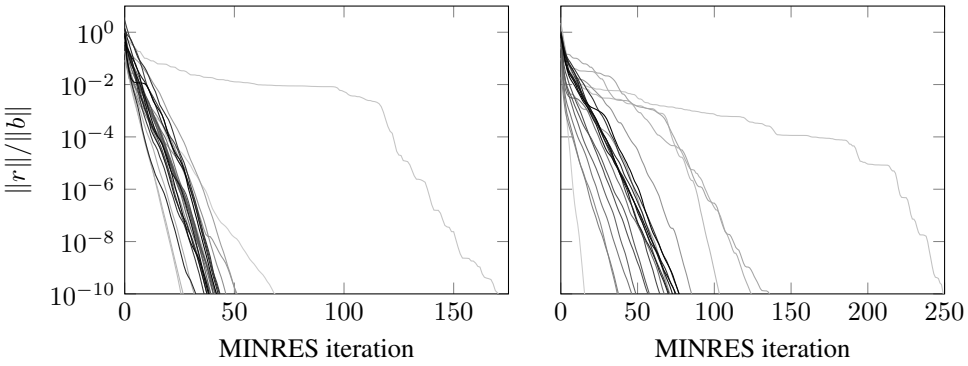
ETNA
Kent State University
http://etna.math.kent.edu



(a) Without deflation.



(b) Deflation of the vector $\mathrm{i}\psi$.



(c) Deflation of 12 Ritz vectors corresponding to the Ritz values of smallest magnitude.

FIG. 3.3. *MINRES convergence histories of all Newton steps for the 2D problem (left) and 3D problem (right). The color of the curve corresponds to the Newton step: light gray is the first Newton step while black is the last Newton step.*

exhibit plateaus of stagnation which are caused by the low-magnitude eigenvalue associated with the near-null space vector $\mathrm{i}\hat{\psi}^{(h)}$.

Figure 3.3b incorporates the deflation of this vector via Algorithm 2.1 with $Y^{(k)} = \mathrm{i}\psi^{(k,h)}$, where $\psi^{(k,h)}$ is the discrete Newton approximate in the kth step. The usage of the preconditioner and the customized inner product (3.11) is crucial here. Clearly, the stagnation effects

ETNA
Kent State University
http://etna.math.kent.edu

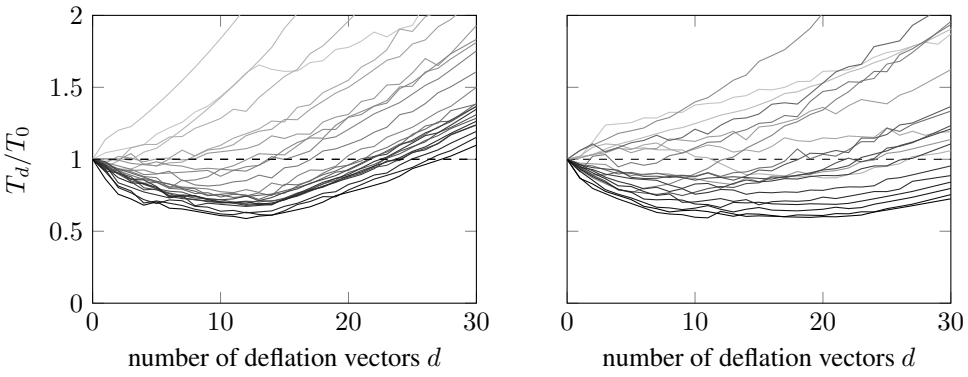


FIG. 3.4. *Wall-times T_d needed for MINRES solves for the test setups (left: 2D; right: 3D) with deflation of those d Ritz vectors from the previous Newton step which correspond to the smallest Ritz values. As in the Figure 3.3, light gray lines correspond to steps early in the Newton process. All times are displayed relative to the computing time T_0 without deflation. The dashed line at $T_d/T_0 = 1$ marks the threshold below where deflation pays off.*

are remedied and a significantly lower number of iterations is necessary to reduce the residual norm to 10^{-10}. While this comes with extra computational cost per step (cf. Table 2.1), this cost is negligible compared to the considerable convergence speedup.

REMARK 3.5. Note that the initial guess $\tilde{x}_0$ is adapted according to (2.14) before the beginning of the iteration. Because of that, the initial relative residual $\|b - Ax_0\|/\|b - A\tilde{x}_0\|$ cannot generally be expected to equal 1 even if $\tilde{x}_0 = 0$. In the particular case of $U = \mathrm{i}\psi$, however, we have

$$x_0 = \mathcal{P}^*\widetilde{x}_0 + U\left\langle U, J(\psi)U\right\rangle_{\mathbb{R}}^{-1}\left\langle U, -\mathcal{S}(\psi)\right\rangle_{\mathbb{R}} = \mathcal{P}^*\widetilde{x}_0$$

since $\langle \mathrm{i}\psi, \mathcal{S}(\psi)\rangle = 0$ by (3.9), and the initial relative residual does equal 1 if $\tilde{x}_0 = 0$ (cf. Figure 3.3b). Note that this is not true anymore when more deflation vectors are added (cf. Figure 3.3c).

Towards the end of the Newton process, a sequence of very similar linear systems needs to be solved. We can hence use the deflated MINRES approach described in Algorithm 2.1, where spectral information is extracted from the previous MINRES iteration and used for deflation in the present process. For the experiments, those 12 Ritz vectors from the MINRES iteration in Newton step k which belong to the Ritz values of smallest magnitude were added for deflation in Newton step $k + 1$. As displayed in Figure 3.3c, the number of necessary Krylov iterations is further decreased roughly by a factor of 2. Note also that in particular the characteristic plateaus corresponding to the low-magnitude eigenvalue do no longer occur. This is particularly interesting since no information about the approximate null space was explicitly specified but automatically extracted from previous Newton steps.

As outlined at the end of Section 2.3, it is a-priori unclear which choice of Ritz-vectors leads to optimal convergence. Out of the choices mentioned in Section 2.3, the smallest-magnitude strategy performed best in the present application.

Technically, one could go ahead and extract even more Ritz vectors for deflation in the next step. However, at some point the extra cost associated with the extraction of the Ritz vectors (Table 2.2) and the application of the projection operator (Table 2.1) will not justify a further increase of the deflation space. The efficiency threshold will be highly dependent on the cost of the preconditioner. Moreover, it is in most situations impossible to predict just how the deflation of a particular set of vectors influences the residual behavior in a Krylov process. For this reason, one has to resort to numerical experiments to estimate the optimal

ETNA
Kent State University
http://etna.math.kent.edu

dimension of the deflation space. Figure 3.4 shows, again for all Newton steps in both setups, the wall clock time of the Krylov iterations as in Figure 3.3 relative to the solution time without deflation. The experiments show that deflation in the first few Newton steps does not accelerate the computing speed. This is due to the fact that the Newton updates are still significantly large and the subsequent linear systems are too different from each other in order to take profit from carrying over spectral information. As the Newton process advances and the updates become smaller, the subsequent linear systems come closer and deflation of a number of vectors becomes profitable. Note, however, that there is a point at which the computational cost of extraction and application of the projection exceeds the gain in Krylov iterations. For the two-dimensional setup, this value is around 12 while in the three-dimensional case, the minimum roughly stretches from 10 to 20 deflated Ritz vectors. In both cases, a reduction of effective computation time by 40% could be achieved.

REMARK 3.6. Other types of deflation vectors can be considered, e.g., harmonic Ritz vectors; see equation (2.20). In numerical experiments with the above test problems we observed that harmonic Ritz vectors resulted in a MINRES convergence behavior similar to regular Ritz vectors. This is in accordance with Paige, Parlett, and van der Vorst [36].

REMARK 3.7. Note that throughout the numerical experiments performed in this paper, the linear systems were solved up to the relative residual of 10^{-10}. In practice, however, one would employ a relaxation scheme as given in, e.g., [10, 39]. Those schemes commonly advocate a relaxed relative tolerance η_k in regions of slow convergence and a more stringent condition when the speed of convergence accelerates toward a solution, e.g.,

$$\eta_k = \gamma \left(\frac{\|F_k\|}{\|F_{k-1}\|} \right)^{\alpha}$$

with some $\gamma > 0$, $\alpha > 1$. In the specific case of nonlinear Schrödinger equations, this means that deflation of the near-null vector $\mathrm{i}\psi^{(k)}$ (cf. Figure 3.3b) becomes ineffective if η_k is larger than the stagnation plateau. The speedup associated with deflation with a number of Ritz vectors (cf. Figure 3.3c), however, is effective throughout the Krylov iteration and would hence not be influenced by a premature abortion of the process.

REMARK 3.8. The numerical experiments in this paper were unavoidably affected by round-off errors. The used MINRES method is based on short recurrences and the sensitivity to round-off errors may be tremendous. Therefore, a brief discussion is provided in this remark. A detailed treatment and historical overview of the effects of finite precision computations on Krylov subspace methods can be found in the book of Liesen and Strakoš [26, Sections 5.8–5.10]. The consequences of round-off errors are manifold and have already been observed and studied in early works on Krylov subspace methods for linear algebraic systems, most notably by Lanczos [25] and Hestenes and Stiefel [22]. A breakthrough was the PhD thesis of Paige [35], where it was shown that the loss of orthogonality of the Lanczos basis coincides with the convergence of certain Ritz values. Convergence may be delayed and the maximal attainable accuracy, e.g., the smallest attainable residual norm, may be way above machine precision and above the user-specified tolerance. Both effects heavily depend on the actual algorithm that is used. In [47] the impact of certain round-off errors on the relative residual was analyzed for an unpreconditioned MINRES variant with the Euclidean inner product. An upper bound on the difference between the exact arithmetic residual r_n and the finite precision residual $\widehat{r}_n$ was given [47, Formula (26)]

$$\frac{\|r_n - \widehat{r}_n\|_2}{\|b\|_2} \leq \varepsilon \left(3\sqrt{3n}\kappa_2(\mathcal{A})^2 + n\sqrt{n}\kappa_2(\mathcal{A}) \right),$$

where ε denotes the machine epsilon. The corresponding bound for GMRES [47, Formula (17)] only involves a factor of $\kappa_2(\mathcal{A})$ instead of its square. The numerical results in [47] also

ETNA
Kent State University
http://etna.math.kent.edu

indicate that the maximal attainable accuracy of MINRES is worse than the one of GMRES. Thus, if very high accuracy is required, the GMRES method should be used. An analysis of the stability of several GMRES algorithms can be found in [5]. In order to keep the finite precision Lanczos basis almost orthogonal, a new Lanczos vector can be reorthogonalized against all previous Lanczos vectors. The numerical results presented in this paper were computed without reorthogonalization, i.e., the standard MINRES method. However, all experiments have also been conducted with reorthogonalization in order to verify that the observed convergence behavior, e.g., the stagnation phases in Figure 3.3a, are not caused by loss of orthogonality.

4. Conclusions. For the solution of a sequence of self-adjoint linear systems such as occurring in Newton process for a large class of nonlinear problems, the authors propose a MINRES scheme that takes into account spectral information from the previous linear systems. Central to the approach is the cheap extraction of Ritz vectors (Section 2.3) out of a MINRES iteration and the application of the projection (2.9).

Differently from similar recycling methods previously suggested [52], the projected operator is self-adjoint and is formulated for inner products other than the ℓ_2-inner product. This allows for the incorporation of a wider range of preconditioners than what was previously possible. One important restriction that is still remaining is the fact that for the orthogonalization of the recycling vectors, the inverse of the preconditioner needs to be known. Unfortunately, this is not the case for some important classes of preconditioners, e.g., multigrid preconditioners with a fixed number of cycles. While this prevents the deflation framework from being universally applicable, the present work extends the range of treatable problems.

One motivating example for this are nonlinear Schrödinger equations (Section 3): the occurring linearization is self-adjoint with respect to a non-Euclidean inner product (3.11), and the computation in a three-dimensional setting is made possible by an AMG preconditioner. The authors could show that for the particular case of the Ginzburg–Landau equations, the deflation strategy reduces the effective run time of a linear solve by up to 40% (cf. Figure 3.3c). Moreover, the deflation strategy was shown to automatically handle the singularity of the problem that otherwise leads to numerical instabilities.

It is expected that the strategy will perform similarly for other nonlinear problems. While adding a number of vectors to the deflation will always lead to a smaller number of Krylov iterations (and thus less applications of the operator and the preconditioner), it only comes with extra computational cost in extracting the Ritz vectors and applying the projection operator; Table 2.2 gives a detailed overview of what entities would need to be balanced. The optimal number of deflated Ritz vectors is highly problem-dependent, in particular dependent upon the computational cost of the preconditioner, and can thus hardly be determined a priori.

The proposed strategy naturally extends to problems which are not self-adjoint by choosing, e.g., GMRES as the hosting Krylov method. For non-self-adjoint problems, however, the effects of altered spectra on the Krylov convergence is far more involved than in the self-adjoint case [32]. This also makes the choice of Ritz vectors for deflation difficult. However, several heuristics for recycling strategies have been successfully applied to non-self-adjoint problems, e.g., by Parks et al. [38], Giraud, Gratton, and Martin [19], Feng, Benner, and Korvink [12] as well as Soodhalter, Szyld, and Xue [49].

Acknowledgments. The authors wish to thank Jörg Liesen for his valuable feedback, Alexander Schlote for providing experimental results with harmonic Ritz vectors and the anonymous referees for their helpful remarks.

ETNA
Kent State University
http://etna.math.kent.edu

REFERENCES

[1] A. R. CHAMPNEYS AND B. SANDSTEDE, *Numerical computation of coherent structures*, in Numerical Continuation Methods for Dynamical Systems, B. Krauskopf, H. M. Osinga, and J. Galán-Vioque, eds., Underst. Complex Syst., Springer, Dordrecht, 2007, pp. 331–358.
[2] A. CHAPMAN AND Y. SAAD, *Deflated and augmented Krylov subspace techniques*, Numer. Linear Algebra Appl., 4 (1997), pp. 43–66.
[3] E. DE STURLER, *Nested Krylov methods based on GCR*, J. Comput. Appl. Math., 67 (1996), pp. 15–41.
[4] Z. DOSTÁL, *Conjugate gradient method with preconditioning by projector*, Int. J. Comput. Math., 23 (1988), pp. 315–323.
[5] J. DRKOŠOVÁ, A. GREENBAUM, M. ROZLOŽNÍK, AND Z. STRAKOŠ, *Numerical stability of GMRES*, BIT, 35 (1995), pp. 309–330.
[6] Q. DU, *Numerical approximations of the Ginzburg-Landau models for superconductivity*, J. Math. Phys., 46 (2005), 095109 (22 pages).
[7] Q. DU, M. D. GUNZBURGER, AND J. S. PETERSON, *Modeling and analysis of a periodic Ginzburg-Landau model for type-*II *superconductors*, SIAM J. Appl. Math., 53 (1993), pp. 689–717.
[8] M. EIERMANN AND O. G. ERNST, *Geometric aspects of the theory of Krylov subspace methods*, Acta Numer., 10 (2001), pp. 251–312.
[9] M. EIERMANN, O. G. ERNST, AND O. SCHNEIDER, *Analysis of acceleration strategies for restarted minimal residual methods*, J. Comput. Appl. Math., 123 (2000), pp. 261–292.
[10] S. C. EISENSTAT AND H. F. WALKER, *Choosing the forcing terms in an inexact Newton method*, SIAM J. Sci. Comput., 17 (1996), pp. 16–32.
[11] H. C. ELMAN, D. J. SILVESTER, AND A. J. WATHEN, *Finite Elements and Fast Iterative Solvers: With Applications in Incompressible Fluid Dynamics*, Oxford University Press, New York, 2005.
[12] L. FENG, P. BENNER, AND J. G. KORVINK, *Subspace recycling accelerates the parametric macro-modeling of MEMS*, Internat. J. Numer. Methods Engrg., 94 (2013), pp. 84–110.
[13] R. W. FREUND, G. H. GOLUB, AND N. M. NACHTIGAL, *Iterative solution of linear systems*, Acta Numer., 1 (1992), pp. 57–100.
[14] A. GAUL, M. H. GUTKNECHT, J. LIESEN, AND R. NABBEN, *A framework for deflated and augmented Krylov subspace methods*, SIAM J. Matrix Anal. Appl., 34 (2013), pp. 495–518.
[15] A. GAUL AND N. SCHLÖMER, *KryPy: Krylov subspace methods package for Python*. https://github.com/andrenarchy/krypy, August 2013.
[16] ———, *PyNosh: Python framework for nonlinear Schrödinger equations*. https://github.com/nschloe/pynosh, August 2013.
[17] M. GEDALIN, T. SCOTT, AND Y. BAND, *Optical solitary waves in the higher order nonlinear Schrödinger equation*, Phys. Rev. Lett., 78 (1997), pp. 448–451.
[18] C. GEUZAINE AND J.-F. REMACLE, *Gmsh: A 3-D finite element mesh generator with built-in pre- and post-processing facilities*, Internat. J. Numer. Methods Engrg., 79 (2009), pp. 1309–1331.
[19] L. GIRAUD, S. GRATTON, AND E. MARTIN, *Incremental spectral preconditioners for sequences of linear systems*, Appl. Numer. Math., 57 (2007), pp. 1164–1180.
[20] A. GREENBAUM, *Iterative Methods for Solving Linear Systems*, SIAM, Philadelphia, 1997.
[21] A. GRIEWANK, *On solving nonlinear equations with simple singularities or nearly singular solutions*, SIAM Rev., 27 (1985), pp. 537–563.
[22] M. R. HESTENES AND E. STIEFEL, *Methods of conjugate gradients for solving linear systems*, J. Research Nat. Bur. Standards, 49 (1952), pp. 409–436.
[23] H. B. KELLER, *The bordering algorithm and path following near singular points of higher nullity*, SIAM J. Sci. Statist. Comput., 4 (1983), pp. 573–582.
[24] M. E. KILMER AND E. DE STURLER, *Recycling subspace information for diffuse optical tomography*, SIAM J. Sci. Comput., 27 (2006), pp. 2140–2166.
[25] C. LANCZOS, *Solution of systems of linear equations by minimized-iterations*, J. Research Nat. Bur. Standards, 49 (1952), pp. 33–53.
[26] J. LIESEN AND Z. STRAKOŠ, *Krylov Subspace Methods: Principles and Analysis*, Oxford University Press, Oxford, 2013.
[27] J. LIESEN AND P. TICHÝ, *Convergence analysis of Krylov subspace methods*, GAMM Mitt. Ges. Angew. Math. Mech., 27 (2004), pp. 153–173.
[28] L. A. M. MELLO, E. DE STURLER, G. H. PAULINO, AND E. C. N. SILVA, *Recycling Krylov subspaces for efficient large-scale electrical impedance tomography*, Comput. Methods Appl. Mech. Engrg., 199 (2010), pp. 3101–3110.
[29] R. B. MORGAN, *A restarted GMRES method augmented with eigenvectors*, SIAM J. Matrix Anal. Appl., 16 (1995), pp. 1154–1171.
[30] ———, *Restarted block-GMRES with deflation of eigenvalues*, Appl. Numer. Math., 54 (2005), pp. 222–236.
[31] R. B. MORGAN AND M. ZENG, *Harmonic projection methods for large non-symmetric eigenvalue problems*,

ETNA
Kent State University
http://etna.math.kent.edu

Numer. Linear Algebra Appl., 5 (1998), pp. 33–55.

[32] M. F. MURPHY, G. H. GOLUB, AND A. J. WATHEN, *A note on preconditioning for indefinite linear systems*, SIAM J. Sci. Comput., 21 (2000), pp. 1969–1972.

[33] R. A. NICOLAIDES, *Deflation of conjugate gradients with applications to boundary value problems*, SIAM J. Numer. Anal., 24 (1987), pp. 355–365.

[34] C. NORE, M. E. BRACHET, AND S. FAUVE, *Numerical study of hydrodynamics using the nonlinear Schrödinger equation*, Phys. D, 65 (1993), pp. 154–162.

[35] C. C. PAIGE, *The computation of eigenvalues and eigenvectors of very large sparse matrices*, PhD. Thesis, Institute of Computer Science, University of London, London, 1971.

[36] C. C. PAIGE, B. N. PARLETT, AND H. A. VAN DER VORST, *Approximate solutions and eigenvalue bounds from Krylov subspaces*, Numer. Linear Algebra Appl., 2 (1995), pp. 115–133.

[37] C. C. PAIGE AND M. A. SAUNDERS, *Solutions of sparse indefinite systems of linear equations*, SIAM J. Numer. Anal., 12 (1975), pp. 617–629.

[38] M. L. PARKS, E. DE STURLER, G. MACKEY, D. D. JOHNSON, AND S. MAITI, *Recycling Krylov subspaces for sequences of linear systems*, SIAM J. Sci. Comput., 28 (2006), pp. 1651–1674.

[39] M. PERNICE AND H. F. WALKER, *NITSOL: a Newton iterative solver for nonlinear systems*, SIAM J. Sci. Comput., 19 (1998), pp. 302–318.

[40] Y. SAAD AND M. H. SCHULTZ, *GMRES: a generalized minimal residual algorithm for solving nonsymmetric linear systems*, SIAM J. Sci. Statist. Comput., 7 (1986), pp. 856–869.

[41] Y. SAAD, M. YEUNG, J. ERHEL, AND F. GUYOMARC'H, *A deflated version of the conjugate gradient algorithm*, SIAM J. Sci. Comput., 21 (2000), pp. 1909–1926.

[42] N. SCHLÖMER, D. AVITABILE, AND W. VANROOSE, *Numerical bifurcation study of superconducting patterns on a square*, SIAM J. Appl. Dyn. Syst., 11 (2012), pp. 447–477.

[43] N. SCHLÖMER AND W. VANROOSE, *An optimal linear solver for the Jacobian system of the extreme type-II Ginzburg-Landau problem*, J. Comput. Phys., 234 (2013), pp. 560–572.

[44] J. R. SHEWCHUK, *Delaunay refinement algorithms for triangular mesh generation*, Comput. Geom., 22 (2002), pp. 21–74.

[45] V. SIMONCINI AND D. B. SZYLD, *Recent computational developments in Krylov subspace methods for linear systems*, Numer. Linear Algebra Appl., 14 (2007), pp. 1–59.

[46] ———, *On the superlinear convergence of MINRES*, in Numerical Mathematics and Advanced Applications 2011, A. Cangiani, R. L. Davidchack, E. Georgoulis, A. N. Gorban, J. Levesley, and M. V. Tretyakov, eds., Springer, Heidelberg, 2013, pp. 733–740.

[47] G. L. G. SLEIJPEN, H. A. VAN DER VORST, AND J. MODERSITZKI, *Differences in the effects of rounding errors in Krylov solvers for symmetric indefinite linear systems*, SIAM J. Matrix Anal. Appl., 22 (2000), pp. 726–751.

[48] B. K. SOM, M. R. GUPTA, AND B. DASGUPTA, *Coupled nonlinear Schrödinger equation for Langmuir and dispersive ion acoustic waves*, Phys. Lett. A, 72 (1979), pp. 111–114.

[49] K. M. SOODHALTER, D. B. SZYLD, AND F. XUE, *Krylov subspace recycling for sequences of shifted linear systems*, Appl. Numer. Math., 81 (2014), pp. 105–118.

[50] G. W. STEWART AND J. G. SUN, *Matrix Perturbation Theory*, Academic Press, Boston, 1990.

[51] J. M. TANG, R. NABBEN, C. VUIK, AND Y. A. ERLANGGA, *Comparison of two-level preconditioners derived from deflation, domain decomposition and multigrid methods*, J. Sci. Comput., 39 (2009), pp. 340–370.

[52] S. WANG, E. DE STURLER, AND G. H. PAULINO, *Large-scale topology optimization using preconditioned Krylov subspace methods with recycling*, Internat. J. Numer. Methods Engrg., 69 (2007), pp. 2441–2468.