BIT

Check for updates

A flexible short recurrence Krylov subspace method for matrices arising in the time integration of port-Hamiltonian systems and ODEs/DAEs with a dissipative Hamiltonian

Malak Diab[1] · Andreas Frommer[1] · Karsten Kahl[1]

Abstract

For several classes of mathematical models that yield linear systems, the splitting of the matrix into its Hermitian and skew Hermitian parts is naturally related to properties of the underlying model. This is particularly so for discretizations of dissipative Hamiltonian ODEs, DAEs and port-Hamiltonian systems where, in addition, the Hermitian part is positive definite or semi-definite. It is then possible to develop short recurrence optimal Krylov subspace methods in which the Hermitian part is used as a preconditioner. In this paper, we develop new, right preconditioned variants of this approach which, as their crucial new feature, allow the systems with the Hermitian part to be solved only approximately in each iteration while keeping the short recurrences. This new class of methods is particularly efficient as it allows, for example, to use few steps of a multigrid solver or a (preconditioned) CG method for the Hermitian part in each iteration. We illustrate this with several numerical experiments for large scale systems.

Keywords Krylov subspace · Short recurrence · Right preconditioning · Optimal methods · Flexible preconditioning · Dissipative Hamiltonian · Port-Hamiltonian systems · Implicit time integration

Mathematics Subject Classification 65F08 · 65F10 · 65L04 · 65L80

✉ Andreas Frommer
frommer@uni-wuppertal.de

Malak Diab
mdiab@uni-wuppertal.de

Karsten Kahl
kkahl@uni-wuppertal.de

[1] Department of Mathematics, Bergische Universität Wuppertal, 42097 Wuppertal, Germany

1 Introduction

Consider the linear system

$$Ax = b, \ \ A \in \mathbb{C}^{n \times n},$$

and the splitting

$$A = H + S$$

with $H = \frac{1}{2}(A + A^*)$ the Hermitian and $S = \frac{1}{2}(A - A^*)$ the skew Hermitian part of A, $H^* = H$, $S^* = -S$. Here, * denotes the adjoint with respect to the standard inner product, i.e., for any matrix B the matrix B^* is obtained by transposing B and taking complex conjugates for each entry.

In several important applications, and in particular in linear systems arising from discretizations of dissipative Hamiltonian ODEs and certain port-Hamiltonian systems, the Hermitian part H is positive definite and "dominates" S in the sense that $H^{-1}S$ or S is small. Then, H is an efficient preconditioner for A, and as we will discuss in Sect. 2, there exist optimal short recurrence Krylov subspace methods using this preconditioner.

Each iterative step of these methods requires one exact solve with the matrix H. The purpose of this paper is to go one step further and investigate practically important short recurrence methods where the solves with H are performed only approximately. For example, we might want to use an incomplete Cholesky factorization in cases where an exact factorization is too expensive, or we might want to perform just some steps of (plain or preconditioned) CG or a multigrid solver for H. We believe that this is an aspect of primordial importance: As long as systems are small enough such that direct factorizations with A are feasible, there is no real need to consider an iterative method which, after all, requires again a direct factorization of a matrix, H, of the same size. Thus we expect that variants which allow for inexact solves of systems with H to have practical impact in the large scale case.

The paper is organized as follows: In Sect. 2 we review the existing methods with exact solves for H and relate them to a known method for systems where the matrix is Hermitian plus a purely imaginary multiple of the identity. In Sect. 3 we then develop the right preconditioned variants of the methods existing in the literature and use those to develop flexible variants. We describe these algorithmically and analyze some of their most relevant properties. In Sect. 5 we present the results of numerical experiments for four examples demonstrating the efficiency gains of the flexible methods over those relying on exact solves.

2 Review of methods with exact solves for H

From now on we assume that the Hermitian part H of A is positive definite. The method proposed by Concus and Golub [3] and, independently, by Widlund [31] as well as the method of Rapoport [21] are, mathematically, equivalent to the full orthog-

onalization method FOM (see [23], e.g.) and the minimal residual method MINRES [20], respectively, for the preconditioned system

$$(I + H^{-1}S)x = H^{-1}b$$

with the standard inner product replaced by the H-inner product $\langle x, y \rangle_H := y^* H x$.

In this section, we present the essentials of both these methods. The starting point is to observe that the matrix $H^{-1}S$ in the preconditioned matrix $H^{-1}A = I + H^{-1}S$ is H-anti-selfadjoint according to the following definition and subsequent proposition (whose proof is straightforward).

Definition 1 A matrix $B \in \mathbb{C}^{n \times n}$ is H-selfadjoint if $\langle Bx, y \rangle_H = \langle x, By \rangle_H$ for all $x, y \in \mathbb{C}^n$. It is H-anti-selfadjoint if $\langle Bx, y \rangle_H = -\langle x, By \rangle_H$ for all $x, y \in \mathbb{C}^n$.

Proposition 1 *A matrix $B \in \mathbb{C}^{n \times n}$ is H-selfadjoint iff*

$$B^* = HBH^{-1}.$$

It is H-anti-selfadjoint if

$$B^* = -HBH^{-1}.$$

Because $I + H^{-1}S$ is a shifted H-anti-selfadjoint matrix, the Arnoldi process for the H-inner product produces a three term recurrence. This is due to the fact that the adjoint of $I + H^{-1}S$ with respect to the H-inner product is a polynomial of degree 1 in $I + H^{-1}S$, a property that was termed "H-normal(1)" in [5]. We refer to this paper and to [14], e.g., for a general discussion on the existence of short recurrences in the Arnoldi process and to [13] for an extension involving inverses.

Introducing the notation $\mathcal{K}_m(A, b)$ for the Krylov subspace $\mathcal{K}_m(A, b) = \text{span}\{b, Ab, \ldots, A^{m-1}b\}$ of order m generated by the matrix A and the vector b, this short recurrence property can be formulated as

$$(I + H^{-1}S)V_m = V_{m+1}T_{m+1,m},$$

where the columns v_i of V_m form an H-orthonormal basis of $\mathcal{K}_m(I + H^{-1}S, b)$ and $T_{m+1,m} \in \mathbb{R}^{(m+1) \times m}$ is tridiagonal (and real). Denoting $T_{m,m}$ the matrix obtained from $T_{m+1,m}$ by removing the last row, we also have that $T_{m,m} - I$ is skew-symmetric. The iterates of the method of Concus and Golub / Widlund are now variationally characterized by a Galerkin condition which we denote ℓGAL to stress the left preconditioning; the iterates of the method of Rapoport are characterized my a minimal residual condition denoted ℓMR. In the description of these variational conditions below, r_0 denotes the initial residual $H^{-1}b - (I + H^{-1}S)x_0$, $\mathcal{K}_m$ denotes $\mathcal{K}_m(I + H^{-1}S, r_0)$.

$$\ell\text{GAL:} \qquad r_m := b - (I + H^{-1}S)x_m \perp_H \mathcal{K}_m, \tag{1}$$

$$\ell\text{MR:} \qquad \|r_m\|_H = \min_{x \in x_0 + \mathcal{K}_m} \|b - (I + H^{-1}S)x\|_H. \tag{2}$$

This gives iterates $x_m = x_0 + V_m \zeta_m$ with

$$\ell\text{GAL:} \qquad T_{m,m}\zeta_m = \|r_0\|_H e_1, \tag{3}$$

$$\ell\text{MR:} \qquad \zeta_m = \operatorname{argmin}_{\zeta \in \mathbb{C}^m} \left\{ \left\| \|r_0\|_H e_1 - T_{m+1,m}\zeta \right\|_2 \right\}. \tag{4}$$

Note that in the formula for the m-th iterate of the ℓMR variant we find the standard 2-norm in the least squares problem for ζ_m.

The short recurrences for computing the iterates of ℓGAL and ℓMR with the H-inner product arise using an LU or a QR factorization of $T_{m,m}$ or $T_{m+1,m}$ which is updated from one iteration to the next. All eigenvalues of $T_{m,m}$ are of the form $1 + i\mu$ with $\mu \in \mathbb{R}$ so that $T_{m,m}$ as well as all its principal minors cannot become singular, which guarantees that a non-pivoted LU-factorization always exists. For the ℓMR variant one typically uses a QR factorization of $T_{m+1,m}$ with an implicit representation of Q as a sequence of Givens rotations and R having three non-zero diagonals (the diagonal and two superdiagonals). Overall this process is analogous to the implementations proposed for SYMMLQ and MINRES in [20], to the implementation of the GAL, MR and ME methods in [6], to the implementation of ℓGAL in [3], [31] and of ℓMR in [21] and to the implementation of QMR as described in [7]. We therefore spare ourselves from reproducing technical details, the principles of which can also be found, e.g., in the textbooks [14], [23] and in [26].

3 Right preconditioned and flexible methods

We now relate the methods of Concus and Golub / Widlund and of Rapoport to the CG type methods considered by Freund in [6] and then develop analogous methods using right preconditioning. This opens the door to allow for an inexact application of H^{-1}, i.e., approximate solutions with H.

If we multiply the original equation $Ax = (H + S)x = b$ with the imaginary unit i we obtain

$$(iH + iS)x = ib. \tag{5}$$

In here $iS =: B$ is Hermitian, and preconditioning with H^{-1} yields a shifted H-selfadjoint matrix $H^{-1}(iH + B) = iI + H^{-1}B$.

In his 1990 paper, Freund [6] considers CG type methods for matrices of the form

$$i\sigma I + W, \quad W^* = W, \sigma \in \mathbb{R}.$$

There, three short recurrence Krylov subspace methods, all within the framework of the standard ℓ_2 inner product, are developed: GAL, MR and ME. GAL is variationally characterized by a Galerkin condition, MR by a minimal residual condition, and ME ("minimal error") minimizes the error in 2-norm over the subspace $x_0 + (i\sigma I + W)^* K_m(i\sigma I + W, r_0)$. The ME approach dates back to [8] for the case $\sigma = 0$ and is less commonly used today. In our numerical experiments it behaved quite similarly as the other methods, but with slightly larger errors and residuals than GAL, so we do not consider this variant any further in this paper.

The methods in [6] can be easily adapted to work with the H-inner product rather than with the standard inner product. The requirement is that W is then an H-selfadjoint matrix, which is exactly the situation that arises when we precondition (5) with H^{-1}, yielding $W = H^{-1}B$ (and $\sigma = 1$). For any vector r, the Krylov subspaces $\mathcal{K}(i\sigma I + H^{-1}B, ir)$ and $\mathcal{K}(\sigma I + H^{-1}S, r)$ are identical. This means that for a given initial guess x_0, the defining Galerkin conditions produce identical iterates, independently of whether we consider $(\sigma I + H^{-1}S)x = b$ with initial residual $r = b - (\sigma I + H^{-1}S)x_0$ or $(i\sigma I + B)x = ib$ with initial residual $ib - (i\sigma I + B)x_0 = ir$. The same holds for the minimal residual based methods. We emphasize this relation in the following proposition for the case $\sigma = 1$.

Proposition 2 *Freund's GAL method with the H-inner product for $(iI + H^{-1}B)x = ib$ is identical to the Concus and Golub / Widlund method for $(I + H^{-1}S)x = b$ in the sense that it produces the same iterates if we have the same initial guess. Freund's MR with the H-inner product for $iI + H^{-1}B$ is, in the same sense, identical to Rapoport's method for $I + H^{-1}S$.*

No method using the H-inner product that is equivalent to Freund's ME seems to have been published so far.

3.1 Right preconditioning

Right preconditioning of $H + S$ with H^{-1} gives rise to the system

$$(I + SH^{-1})\hat{x} = b, \ \ x = H^{-1}\hat{x}.$$

As with left preconditioning, there is a direct and simple correspondence with the "Hermitian" formulation on which we will focus in what follows, i.e.

$$(i\sigma I + BH^{-1})\hat{x} = ib, \ \text{with} \ B = iS \ \text{Hermitian} \ \text{and} \ \sigma \in \mathbb{R}.$$

Herein, $W = BH^{-1}$ is H^{-1}-selfadjoint, and we again obtain short recurrence methods GAL, MR, ME by using the H^{-1}-inner product in the methods and algorithms of [6].

Proposition 3 *There exist short recurrence GAL, MR and ME methods for the right preconditioned system $i\sigma I + BH^{-1}$. These give rise to short recurrence methods for $i\sigma H + B$ when using the H^{-1} inner product.*

Although it might not be obvious at first sight, the implementation of right preconditioning can be done by investing one multiplication with B, one with H and just one solve with H per iteration. The idea is to carry an additional vector $\tilde{v}$ which stores $H^{-1}v$ for the current Lanczos vector v. Details are given in Algorithm 1 in which, when solving the linear system (5), we take the initial vector w as the initial residual $ib - (i\sigma H + B)x_0$. We note that an implementation with similar cost is possible for left preconditioning. Also note that in Algorithm 1 we

have

$$
\begin{aligned}
\langle w, \widetilde{v}_{k-1} \rangle &= \langle (i\sigma I + BH^{-1})v_k, H^{-1}v_{k-1} \rangle \\
&= \langle H^{-1}v_k, (-i\sigma I + BH^{-1})v_{k-1} \rangle \\
&= \langle H^{-1}v_k, (i\sigma I + BH^{-1})v_{k-1} \rangle \quad (\text{since } \langle v_k, v_{k-1} \rangle_{H^{-1}} = 0) \\
&= \langle H^{-1}v_k, \beta_{k-1}v_k - \alpha_{k-1}v_{k-1} - \beta_{k-2}v_{k-2} \rangle \\
&= \beta_{k-1},
\end{aligned}
$$

which is why we can use β_{k-1} in the H^{-1}-orthogonalization of w against v_{k-1} in line 6.

Algorithm 1 m steps of Lanczos for $i\sigma I + BH^{-1}$ with the H^{-1}-inner product. $\langle \cdot, \cdot \rangle$ denotes the standard inner product

---

1: choose initial vector w, compute $\widetilde{w} = H^{-1}w$, $\beta_0 = \langle w, \widetilde{w} \rangle$
2: put $v_1 = w/\beta_0$, $\widetilde{v}_1 = \widetilde{w}/\beta_0$, $v_0 = 0$
3: **for** $k = 1, \ldots, m$ **do**
4:   $w = i\sigma v_k + B\widetilde{v}_k$ ▷ we have $\widetilde{v}_k = H^{-1}v_k$ from the previous iteration
5:   $\alpha_k = \langle w, \widetilde{v}_k \rangle$
6:   $w = w - \alpha_k v_k - \beta_{k-1}v_{k-1}$
7:   $\widetilde{w} = H^{-1}w$
8:   $\beta_k = \langle w, \widetilde{w} \rangle^{1/2}$
9:   $v_{k+1} = w/\beta_k$
10:   $\widetilde{v}_{k+1} = \widetilde{w}/\beta_k$ ▷ now $\widetilde{v}_{k+1} = H^{-1}v_{k+1}$
11: **end for**

One can work with $\sigma I + SH^{-1}$ rather than with $i\sigma I + BH^{-1}$ in Algorithm 1, adapting the matrix vector multiplication in line 4 and changing the last summand $-\beta_{k-1}v_{k-1}$ to $+\beta_{k-1}v_{k-1}$ in line 6. The reason for the latter is that due to the fact that SH^{-1} is H^{-1}-anti selfadjoint we now have $\langle w, \widetilde{v}_{k-1} \rangle = -\langle H^{-1}v_k, \beta_{k-1}v_k \rangle$. This observation represents the general transition rule from formulations for the systems with $i\sigma I + BH^{-1}$, $i\sigma I + H^{-1}B$ or $i\sigma H + B$ to systems with $\sigma I + SH^{-1}$, $\sigma I + H^{-1}S$ or $\sigma H + S$. Our presentation will continue to be based on the systems using $B = iS$ rather than S.

Right preconditioning is the key to develop flexible variants of the GAL, MR and ME methods. The term "flexible" includes "inexact" methods in which $\widetilde{w} = H^{-1}w$ is replaced by $\widehat{w} = \widehat{H}^{-1}w$ with a fixed matrix $\widehat{H}$. For example, $\widehat{H}$ may result from an incomplete Cholesky factorization. "Flexible" also includes the more general possibility to approximate $H^{-1}w$ via a non-stationary iteration such as (possibly preconditioned) CG which, formally, means that $\widehat{H}$ changes from one approximate solve to the next.

3.2 Flexible methods

Using exact multiplications with H^{-1}, the Lanczos algorithm given in Algorithm 1 for the right preconditioned matrix $i\sigma I + BH^{-1}$ produces the relation

$$(i\sigma I + BH^{-1})V_m = V_{m+1}T_{m+1,m}, \tag{6}$$

with

$$T_{m+1,m} = \begin{bmatrix} \alpha_1 & \beta_1 & & & \\ \beta_1 & \alpha_2 & \beta_2 & & \\ & \beta_2 & \ddots & \ddots & \\ & & \ddots & \ddots & \beta_{m-1} \\ & & & \beta_{m-1} & \alpha_m \\ & & & & \beta_m \end{bmatrix},$$

where now V_m has H^{-1}-orthonormal columns. The transformation of an iterate

$$\widetilde{x}_m = x_0 + V_m \zeta_m$$

for the right preconditioned system back to the iterate $x_m = H^{-1}\widetilde{x}_m = H^{-1}x_0 + H^{-1}V_m\zeta_m$ of the original system is most easily done via a short recurrence update using the preconditioned Lanczos vectors $\widetilde{v}_m = H^{-1}v_m$ which we compute anyways. The relation (6) can then be stated as

$$(i\sigma H + B)\underbrace{\widetilde{V}_m}_{=H^{-1}V_m} = V_{m+1}T_{m+1,m}, \tag{7}$$

with the upper square $m \times m$ block $T_{m,m}$ of $T_{m+1,m}$ being symmetric. We purposefully formulated Algorithm 1 using classical Gram-Schmidt instead of modified Gram-Schmidt in the orthogonalization, since this is what we will need in its flexible variant.

We can now devise the flexible variant in a manner similar to what has been done for FGMRES, the flexible GMRES method [30] and, with regard to short recurrences, in flexible QMR [27], e.g. Assume that z_k is only an approximation to $H^{-1}v_k$ and proceed as follows:

1. Compute $w = (i\sigma H + B)z_k$
2. H^{-1}-orthogonalize w against v_k and v_{k-1} *approximately*. This means that the inner products $\langle w, v_k \rangle_{H^{-1}}$ and $\langle w, v_{k-1} \rangle_{H^{-1}}$ are obtained as standard inner products with $\widehat{w}$ where $\widehat{w}$ is an approximation to $H^{-1}w$.
3. Approximately H^{-1}-normalize the vector resulting from the approximate H^{-1}-orthogonalization.

Algorithm 2 gives the details for one iteration of the resulting right preconditioned flexible Lanczos process.

Algorithm 2 m steps of right preconditioned flexible Lanczos for $i\sigma H + B$ with approximate solves for H. $\langle \cdot, \cdot \rangle$ denotes the standard inner product

---

1: choose initial vector w, compute $\widehat{w} \approx H^{-1}w$, $\beta_0 = \langle w, \widehat{w} \rangle$
2: put $v_1 = w/\beta_0$, $z_1 = \widehat{w}/\beta_0$, $v_0 = 0$
3: **for** $k = 1, \ldots, m$ **do**
4:     $w = (i\sigma H + S)z_k$ ▷ we have $z_k \approx H^{-1}v_k$ from the previous iteration
5:     $\gamma_k = \langle w, z_{k-1} \rangle$
6:     $\alpha_k = \langle w, z_k \rangle$
7:     $w = w - \alpha_k v_k - \gamma_k v_{k-1}$
8:     $\widehat{w} \approx H^{-1}w$
9:     $\beta_k = \langle w, \widehat{w} \rangle^{1/2}$
10:    $v_{k+1} = w/\beta_k$
11:    $z_{k+1} = \widehat{w}/\beta_k$ ▷ now $z_{k+1} \approx H^{-1}v_{k+1}$
12: **end for**

In the non-flexible Algorithm 1 we used $\langle w, \widetilde{v}_{k-1} \rangle = \langle H^{-1}v_k, v_k \rangle^{1/2} = \beta_{k-1}$, and we avoided to compute $\langle w, \widetilde{v}_{k-1} \rangle$ explicitly. In the flexible context the above equality does no longer hold, since $\langle w, \widetilde{v}_{k-1} \rangle$ is computed with the approximation z_{k-1} for $H^{-1}v_{k-1}$, whereas β_k is computed with the approximation $\widehat{w}$ for $H^{-1}w$.

The lack of exact H^{-1}-orthogonality between v_{k-1} and v_k is the reason why we have $\alpha_k \neq \langle w - \gamma_k v_{k-1}, z_k \rangle$ in Algorithm 2, showing that modified Gram-Schmidt orthogonalization is not an option with inexact inner products. Modified Gram-Schmidt is known to numerically preserve orthogonality better than classical Gram-Schmidt for exact inner products, but this property becomes less significant in the flexible case where the H^{-1}-inner products are not zero anyways.

With the flexible Lanczos process we have the recurrence relation

$$(i\sigma H + B)Z_m = V_{m+1}T_{m+1,m}, \tag{8}$$

where $Z_m = [z_1| \cdots |z_m]$ and $T_{m+1,m}$ is again tridiagonal,

$$T_{m+1,m} = \begin{bmatrix} \alpha_1 & \gamma_2 & & & \\ \beta_1 & \alpha_2 & \gamma_3 & & \\ & \beta_2 & \ddots & \ddots & \\ & & \ddots & \ddots & \gamma_m \\ & & & \beta_{m-1} & \alpha_m \\ & & & & \beta_m \end{bmatrix}. \tag{9}$$

Note that $T_{m,m}$ is no longer symmetric. When used to solve the linear system (5), the initial vector w in Algorithm 2 is the residual $ib - (i\sigma H + B)x_0$, i.e., the first Lanczos vector v_1 is

$$v_1 = \frac{1}{\beta_0}r_0 \text{ with } r_0 = ib - (iH + B)x_0, \ \widehat{w} \approx H^{-1}r_0, \ \beta_0 = \langle r_0, \widehat{w} \rangle^{1/2}. \tag{10}$$

We emphasize that V_m is now only approximately H^{-1}-orthonormal and Z_m is only approximately equal to $H^{-1}V_m$. We define the flexible GAL and MR iterates in analogy to (3) and (4) as $x_m = x_0 + Z_m\zeta_m$ with

$$\text{FGAL:} \qquad\qquad T_{m,m}\zeta_m = \beta_0 e_1, \qquad\qquad (11)$$

$$\text{FMR:} \qquad\qquad \zeta_m = \text{argmin}_{\zeta \in \mathbb{C}^m}\{\|\beta_0 e_1 - T_{m+1,m}\zeta\|_2\}, \qquad (12)$$

where now V_m, $T_{m+1,m}$ and $T_{m,m}$ obey relation (8). Computationally, we obtain short recurrences for the iterates x_m by updating factorizations of $T_{m,m}$ and $T_{m+1,m}$ just as in the non-flexible case; see the discussion at the end of Sect. 2. Commented Matlab implementations of FMR and FGAL are available from a Gitlab server; see the footnote in Sect. 5.

Note that the FGAL iterates do not fulfill a strict H^{-1}-orthogonality relation for the residuals, nor do the FMR iterates minimize the H^{-1}-norm of the residual exactly. But, of course, if we solve for H exactly in the flexible Lanczos process, FGAL and FMR reduce to right preconditioned counterparts of ℓFOM and ℓMR from (3) and (4), respectively.

A flexibly preconditioned CG method (FCG) for Hermitian positive definite matrices using a preconditioner which may change from one iteration to the next has been proposed in [9] and further analyzed in [19]. The FGAL method can be used in this situation, too, with $\sigma = 0$. FCG and FGAL are similar in spirit but not identical: FCG constructs search directions by applying the variable preconditioner to the current residual and then A-orthogonalizing against a fixed number s of the previous search directions. In contrast, FGAL orthogonalizes the next preconditioned Lanczos vector in the ℓ_2-inner product against the two previous ones. This is different, even when $s = 2$.

4 Properties

In this section we first state two known convergence bounds for the left preconditioned methods where systems with H are solved exactly. We then show that a result from [6] can be used to improve one of these bounds, a fact that seems to have gone unnoticed so far. We continue the section shortly considering the right preconditioned methods and then turn to the flexible variants for which we establish two somewhat more basic but algorithmically important properties.

Theorem 1 *Let $\lambda \geq 0$ be the smallest value such that the interval $[-i\lambda, i\lambda]$ contains all the (purely imaginary) eigenvalues of the H-anti-selfadjoint matrix $H^{-1}S$ and let x_* denote the solution of the system $Ax = b$. Then*

(i) *For the ℓGAL iterates x_m from (1) we have*

$$\|x_{2m} - x_*\|_H \leq \|x_0 - x_*\|_H \cdot 2\left(\frac{\sqrt{1+\lambda^2}-1}{\sqrt{1+\lambda^2}+1}\right)^m$$

and

$$\|x_{2m+1} - x_*\|_H \leq \|x_1 - x_*\|_H \cdot 2 \left(\frac{\sqrt{1 + \lambda^2} - 1}{\sqrt{1 + \lambda^2} + 1} \right)^m.$$

(ii) *For the ℓMR residuals $r_m = b - Ax_m$ from* (2) *we have*

$$\|b - Ax_m\|_{H^{-1}} \leq \|b - Ax_0\|_{H^{-1}} \cdot 2 \left(\frac{\lambda}{\sqrt{1 + \lambda^2} + 1} \right)^m.$$

Part (i) of the theorem goes back to [4, 12, 28], and part (ii) to [28]; see [10] for a detailed discussion. What seems to have gone unnoticed so far is that a better bound for the ℓMR iterates results from [6, Theorem 4], adapted to H-selfadjoint matrices. We first formulate the result for an arbitrary value for the shift σ before comparing to Theorem 1(ii). Note that the original theorem in [6] considers residuals for the matrix $i\sigma I + W$ with W Hermitian and the ℓ_2-norm. For the residuals belonging to matrices $i\sigma I + W$ with W H-selfadjoint it holds in exactly the same manner, but now with the H-inner product and associated norm. When translating back to our original matrix $H + S$, we have $\sigma = 1$, and H-norms for the residual w.r.t. $iI + H^{-1}B$ turn into H^{-1}-norms for the residuals w.r.t. $H + S$. This gives us the following theorem.

Theorem 2 *Let* $\mathrm{spec}(H^{-1}S) \subseteq i[\alpha, \beta]$ *with* $\alpha < \beta$ *and let R be the unique solution of*

$$\frac{1}{2} \left(R + \frac{1}{R} \right) = \frac{\sqrt{\beta^2 + 1} + \sqrt{\alpha^2 + 1}}{\beta - \alpha}, \quad R > 1. \tag{13}$$

Then the residuals of the ℓMR iterates x_m from (2) *satisfy*

$$\|b - Ax_m\|_{H^{-1}} \leq \|b - Ax_0\|_{H^{-1}} \cdot \frac{2}{R^k + R^{-k}}.$$

Compared to Theorem 1 this theorem allows for a more narrow interval to contain the spectrum of $H^{-1}S$ as it does not need to be symmetric with respect to the real axis. Note, however, that for real matrices $\mathrm{spec}(H^{-1}S)$ is always symmetric with respect to the real axis, so that optimal bounds on the spectrum are of the form $i[-\lambda, \lambda]$ as in Theorem 1. But even with symmetric bounds for the spectrum, the bound of Theorem 2 is always better than that of Theorem 1 as we will explain now. In order to not unnecessarily burden our discussion, we will omit details of somewhat longish but elementary algebraic manipulations in the paragraph to follow.

We first note that the expression $(\sqrt{\beta^2 + 1} + \sqrt{\alpha^2 + 1})/(\beta - \alpha)$ increases monotonically in β and decreases monotonically in α. Thus, for $\lambda \geq \max\{\beta, -\alpha\}$, we have

$$\frac{\sqrt{\beta^2 + 1} + \sqrt{\alpha^2 + 1}}{\beta - \alpha} \leq \frac{\sqrt{\lambda^2 + 1}}{\lambda}.$$

The solution $R > 1$ of the equation $R + 1/R = c$ with $c \geq 1$ increases monotonically with c. Consequently, when $\lambda \geq \max\{\beta, -\alpha\}$, the solution R of (13) is larger or equal than the solution $\widehat{R} = \left(\sqrt{1 + \lambda^2} + 1\right)/\lambda$ of $R + 1/R = \sqrt{\lambda^2 + 1}/\lambda$. Now, the factor $\left(\lambda/(\sqrt{1 + \lambda^2} + 1)\right)^m$ in the bound of Theorem 1(ii) is precisely $1/\widehat{R}^m$. The quantity $1/(R^m + R^{-m})$ in the bound of Theorem 2 is thus smaller for two reasons: Because of the presence of $R^{-m} > 0$ and because $\widehat{R} > R$ as soon as $\beta \neq -\alpha$, i.e., $\lambda > \beta$ or $\lambda > -\alpha$.

Note, that Theorem 2 is also applicable for the exact right preconditioned MR variant as introduced at the beginning of Sect. 3.1. Analogously to the derivation in Theorem 2, the results for the MR method from [6] can be applied to the matrix $iI + BH^{-1}$—in which BH^{-1} is now H^{-1}-selfadjoint—and the H^{-1}-inner product. Accordingly, the residuals with respect to $iI + BH^{-1}$ and $iH + B$ are now identical since $b - (iI + BH^{-1})\tilde{x}_m = b - (iH + B)x_m$ with $H^{-1}\tilde{x}_m = x_m$, so that the bounds hold for the H^{-1}-norm, in analogy to the left preconditioned case. Further note that $\mathrm{spec}(SH^{-1}) = \mathrm{spec}(H^{-1}S)$, which means that we can use the same bounds α and β as in the left preconditioned case. In a similar manner, the bounds in Theorem 1(i) carry over to the right preconditioned case: We first obtain bounds for the H^{-1}-norm of the preconditioned iterates $\hat{x}_m$, and transforming back to the iterates $x_m = H^{-1}\hat{x}_m$ of the original system just changes the H^{-1}-norm to the H-norm.

We now consider the flexible methods. Conceivably, a detailed convergence analysis, depending on the "degree of inexactness" of the preconditioner could be based on prior work for flexible CG [19], flexible GMRES [22, 30] and, in particular, flexible QMR [27]. We anticipate this to be quite involved and also quite technical, so that we do not address this in the context of the present paper. We rather state and prove two basic properties which are essential for the algorithmic aspects.

The first property is that in typical situations we cannot encounter unlucky breakdowns in the flexible Lanczos process, Algorithm 2. An unlucky breakdown at step m means that we get $\beta_m = \langle w, \widehat{w} \rangle = 0$ although $w \neq 0$. If there is no unlucky breakdown, the FMR iterates all exist up to the step where we obtain the solution, as in that case the matrix $T_{m+1,m}$ of (9) has full rank for all relevant m. Though, its square part $T_{m,m}$ need not be non-singular, which means that the FGAL iterate does not necessarily exist. However, as is done in the SYMMLQ-algorithm (see [14, 20]), the QR-factorization of $T_{m,m}$ can still be updated from that of $T_{m-1,m-1}$, which allows to efficiently update the next FGAL iterate from the previous existing one using the QR-factorization.

Proposition 4 *Assume that the inexact solves $\widehat{w} \approx H^{-1}w$ in Algorithm 2 are done in one of the following ways*

(i) *by performing k steps of the CG method with initial guess 0,*
(ii) *by performing k steps of the preconditioned CG method with initial guess 0 and a Hermitian positive definite preconditioner K,*
(iii) *as $\widehat{w} = Gw$ with $G \in \mathbb{C}^{n \times n}$ Hermitian and positive definite.*

Then, if $w \neq 0$, we have $\beta = \langle w, \widehat{w} \rangle > 0$.

Proof In case (i), the initial residual for the CG method is w, and with the columns of $W \in \mathbb{C}^{n \times k}$ being the Lanczos vectors which form an orthonormal basis of the Krylov subspace $\mathcal{K}_k(H, w)$, we know that $\widehat{w} = V_k(V_k^* H V_k)^{-1} V_k^* w$; see e.g. [14] or [23]. Thus, $\beta_m = \langle w, \widehat{w} \rangle = \langle V_k^* w, (V_k^* H V_k)^{-1} V_k^* w \rangle$, and this quantity is positive as $V_k^* H V_k$ is Hermitian positive definite and $V_k^* w = \|w\|_2 e_1$, e_1 the first canonical unit vector in $\mathbb{C}^k$, is non-zero.

Case (ii) follows in a similar manner as (i), replacing V_k by the matrix of K-orthonormal Lanczos vectors for $\mathcal{K}_k(K^{-1}H, K^{-1}w)$.

In case (iii), as G is Hermitian positive definite and $w \neq 0$, we immediately have $\beta = \langle w, \widehat{w} \rangle = \langle w, Gw \rangle > 0$.

An important example for case (iii) above is when $G = L^{-1}L^{-*}$ with L a sparse lower triangular matrix arising from an incomplete Cholesky factorization of H, $H = L^*L - R$. Such G is also a typical preconditioner K^{-1} for case (ii). Another important example for (iii) are Galerkin based multigrid methods for Hermitian positive definite systems with symmetric pre- and post-smoothing and exact coarsest level solves.

The second property we want to stress deals with $\|b - Ax_m\|_{H^{-1}}$ as a measure of the error of iterate x_m. Using $b - Ax_m = A(x_* - x_m)$, we have

$$
\begin{aligned}
\|b - Ax_m\|_{H^{-1}}^2 &= \langle A(x_* - x_m), A(x_* - x_m) \rangle_{H^{-1}} \\
&= \langle x_* - x_m, x_* - x_m \rangle_{A^* H^{-1} A} = \|x_* - x_m\|_{A^* H^{-1} A}^2
\end{aligned}
$$

with

$$
A^* H^{-1} A = (-i\sigma H + B) H^{-1} (i\sigma H + B) = \sigma^2 H + B H^{-1} B.
$$

Thus, if H dominates $B = iS$, i.e., if $H^{-1}B$ is small, the matrix $AH^{-1}A$ is close to $\sigma^2 H$ (and we have equality if $B = 0$). Consequently, when H dominates B, up to the trivial factor σ^2, the norm $\|b - Ax_m\|_{H^{-1}}$ is a good approximation to the H-norm of the error. There are multiple reasons why in the case $B = 0$ the H-norm of the error—and not its ℓ_2-norm nor the ℓ_2-norm of the residual, e.g.—should be considered the most adequate measure for the error, in particular when the linear systems arise from a discretization of an underlying continuous, infinite-dimensional equation; see e.g. [1, 15]. Even if $B \neq 0$, the H-norm can still be regarded as the canonical measure for the error. By the following proposition, an approximate bound for this norm is available at almost no cost for the FMR iterates.

Proposition 5 *Let* v_k, $k = 1, \ldots, m$, *denote the vectors produced by the flexible Lanczos algorithm, Algorithm 2, and let* ϱ_m *be the value of the minimum in the defining equation* (12) *of the FMR iterate* x_m, *i.e.,*

$$
\varrho_m := \min_{\zeta \in \mathbb{C}^m} \left\{ \|\beta_0 e_1 - T_{m+1,m} \zeta \|_2 \right\}.
$$

Then

$$
\|b - Ax_m\|_{H^{-1}} \leq \|V_{m+1}\|_{H^{-1}} \cdot \varrho_m, \tag{14}
$$

and, if the flexible Lanczos vectors z_m approximate the "exactly inverted" vectors $H^{-1}v_m$ to a relative accuracy of ε in the H-norm, i.e.,

$$\|z_m - H^{-1}v_m\|_H \leq \varepsilon \cdot \|H^{-1}v_m\|_H, \tag{15}$$

then

$$\|V_{m+1}\|_{H^{-1}} \leq \left(\frac{m+1}{1-\varepsilon}\right)^{1/2}. \tag{16}$$

Proof Denote by $V_m = [v_1 \mid \cdots \mid v_m] \in \mathbb{C}^{n \times m}$ the matrix which contains the flexible Lanczos vectors v_k as its columns. We have $x_m = x_0 + V_m \zeta_m$ with ζ_m satisfying the minimality condition (12). Thus, using the flexible Lanczos relation (8) and (10) we obtain

$$
\begin{aligned}
\|b - Ax_m\|_{H^{-1}} &= \|V_{m+1}\beta_0 e_1 - V_{m+1}T_{m+1,m}\zeta_m\|_{H^{-1}} \\
&= \|H^{-1/2}V_{m+1}(\beta_0 e_1 - T_{m+1,m}\zeta_m)\|_2 \\
&\leq \|H^{-1/2}V_{m+1}\|_2 \cdot \|\beta_0 e_1 - T_{m+1,m}\zeta_m\|_2 \\
&= \|H^{-1/2}V_{m+1}\|_2 \cdot \varrho_m,
\end{aligned}
$$

which is (14). To obtain (16) we now show that the ℓ_2-norm of each column of $H^{-1/2}V_{m+1}$ is bounded by $\sqrt{1/(1-\varepsilon)}$. The inequality (16) then follows as the ℓ_2-norm of a matrix never exceeds its Frobenius norm.

With $\nu_k := \|v_k\|_{H^{-1}} = \|H^{-1/2}v_k\|_2$ we have

$$\nu_k^2 = \langle v_k, z_k\rangle + \langle v_k, H^{-1}v_k - z_k\rangle = 1 + \langle H^{-1/2}v_k, H^{-1/2}v_k - H^{1/2}z_k\rangle,$$

from which we get, via the Cauchy-Schwarz inequality,

$$|\nu_k^2 - 1| \leq \|H^{-1/2}v_k\| \cdot \|H^{-1/2}v_k - H^{1/2}z_k\| = \|v_k\|_{H^{-1}} \cdot \|H^{-1}v_k - z_k\|_H.$$

With assumption (15) we therefore have

$$|\nu_k^2 - 1| \leq \varepsilon \nu_k^2,$$

which is equivalent to

$$\frac{1}{1+\varepsilon} \leq \nu_k^2 \leq \frac{1}{1-\varepsilon}.$$

There are two important observations regarding Proposition 5. First, practically, we cannot monitor H-norms of the error, so inequality (15) cannot be checked directly. However, a relative reduction of the (ℓ_2-norm of the) residual, which can be measured, typically results in a similar reduction of the error. In case that we use a multigrid method to solve systems with H, this choice is satisfied since the residual 2-norm and the H-norm of the error are of the same order of magnitude. This is due to the fact that multigrid methods attempt to uniformly reduce all components in the

spectral decomposition of the error; see [29]. In the case that we use the CG method to approximate $H^{-1}v_m$, it is precisely the H-norm of the error which is made as small as possible among all vectors from the k-th affine Krylov subspace by the k-th CG-iterate $z_m^{(k)}$. This norm, $\|z_m^{(k)} - H^{-1}v_m\|_H$, can be estimated accurately using quantities available from CG iterations $k, k+1, \ldots, k+d$, and the recent paper [18] gives a particularly efficient implementation in which the best value for d is determined adaptively. Note that in this case $z_m^{(k+d)}$ has already been computed, and this $(k+d)$-th iterate has an even smaller H-norm of the error than the k-th iterate $z_m^{(k)}$. Thus the price to pay for accurate estimates of the error norm are additional CG iterations. We refer to [24] for a further discussion of this and related aspects.

Second, the quantity ϱ_m is available in the algorithm at no additional cost. The situation is similar as in GMRES [23] or QMR [7]: To solve the least squares problem (12), we update the QR-factorization $T_{m+1,m} = Q_{m+1,m+1}R_{m+1,m}$ in each step. The value ϱ_m is the last, i.e., the $(m+1)$-st, entry of $Q_{m+1,m}^*\|r_0\|_H e_1$, and this value is updated to ϱ_{m+1} using the Givens rotation which updates $Q_{m+1,m}$ to $Q_{m+2,m+1}$.

We further note that the proof of Proposition 5 also shows that in the case where we solve for H exactly, we find $\varrho_m = \|b - Ax_m\|_{H^{-1}}$ since $\|V_m\|_{H^{-1}} = 1$.

5 Numerical results

In this section we report results for several examples which have been considered in the literature. The general setup for all experiments is as follows: We report residual norm plots for both, the FMR and the FGAL method. In each iteration, the systems with the Hermitian positive definite part H are solved approximately using the CG method, asking for the initial ℓ_2-norm of the residual to be reduced by a factor of ε_{cg}. We typically vary ε_{cg} from 10^{-1} up to 10^{-12}. In light of Proposition 5 and the discussion preceding it, the FGAL and FMR iteration is stopped once the H^{-1}-norm of the initial residual is reduced by $\varepsilon_f = 10^{-12}$. The case $\varepsilon_{cg} = 10^{-12}$ in the approximate solves may thus actually be considered as equivalent to exact solves in our context.

To facilitate the reproducibility of our tests, we published the Matlab implementation of the methods we use along with the test cases we study on GitLab.[1] The code in this repository includes an implementation of CG from [18], which stops CG at iteration $k+d$ when a prescribed reduction of the accurately estimated H-norm of the error of the iterate at iteration k is obtained. The residual norm plots for our examples report (relative) approximate H^{-1}-norms of the residual, i.e., the quantity $\sqrt{\widehat{r}^*r}$ where r is the current residual and $\widehat{r}$ is the result of CG when solving $H\widehat{r} = r$ with initial guess 0 and a required reduction of the residual ℓ_2-norm of ε_{cg}. This number can be somewhat off the exact H^{-1}-norm; the advantage is that it is computed with an effort comparable to what is needed to update an iterate. When the iterations stopped, we double-checked that the targeted decrease of the H^{-1}-norm was indeed achieved by recomputing $\sqrt{\widehat{r}^*r}$ for the final residual with now $\varepsilon_{cg} = 10^{-12}$ in the CG method which computes $\widehat{r}$. Finally, as an illustration, our plots also give the bound on the H^{-1}-norm of the residuals of the FMR iterates according to Proposition 5. Let us
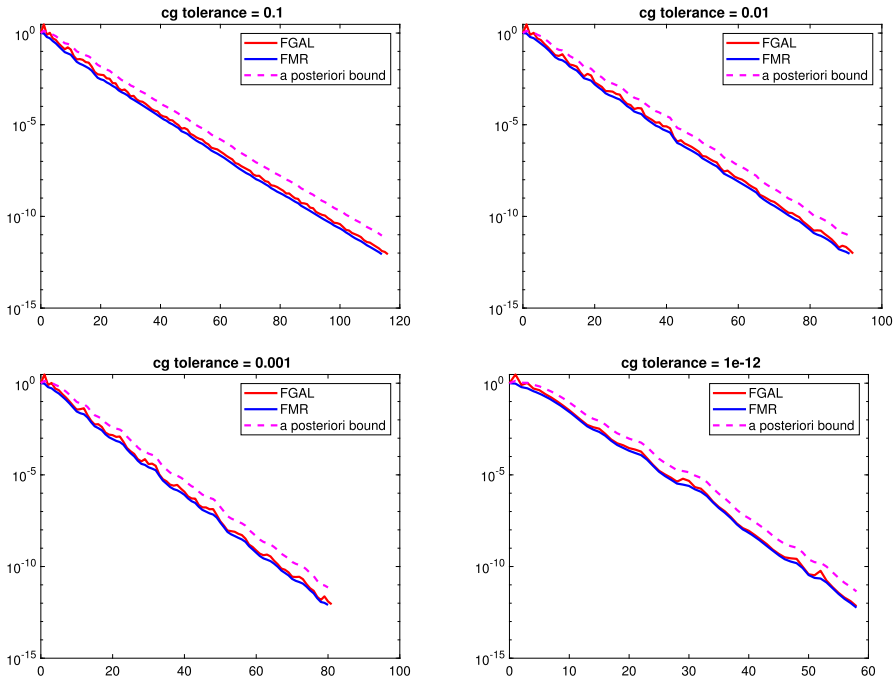
---

[1] Available at https://git.uni-wuppertal.de/anginf/krylov4phs.

Fig. 1 Convergence plots for the convection–diffusion equation, Example 1, four choices for ε_{cg}

stress that, as opposed to the approximate H^{-1}-norms, these bounds can be computed at virtually no extra cost, so that in a production setting we could base a stopping criterion for the FMR iterations merely on these bounds.

We start reconsidering the example in [31].

Example 1 Consider the convection–diffusion equation

$$-\Delta u(x, y) + a u_x(x, y) = b(x, y),$$

on $\Omega = [0, 1] \times [0, 1]$ with Dirichlet boundary conditions. We discretize on a uniform square mesh with 127×127 interior grid points, using central finite differences in both, Δu and u_x, and choose $a = 10^4$. The right hand side b was chosen to have random entries.

Figure 1 shows convergence plots for four different choices of ε_{cg}. We see that for all ε_{cg} the residual norms of the FGAL and the FMR iterates differ only marginally, with the FMR residuals, as is to be expected due to their (approximate) minimization property, being (slightly) smaller and decreasing in a smoother manner. We also see that the bound from Proposition 5 captures the actual convergence behavior quite accurately. With $\varepsilon_{cg} = 10^{-1}$, we already need only about twice as many iterations as with "exact" solves, $\varepsilon_{cg} = 10^{-12}$, while an approximate solve with $\varepsilon_{cg} = 10^{-1}$ requires about 50 CG iterations on average as opposed to about 470 CG iterations on average for $\varepsilon_{cg} = 10^{-12}$.
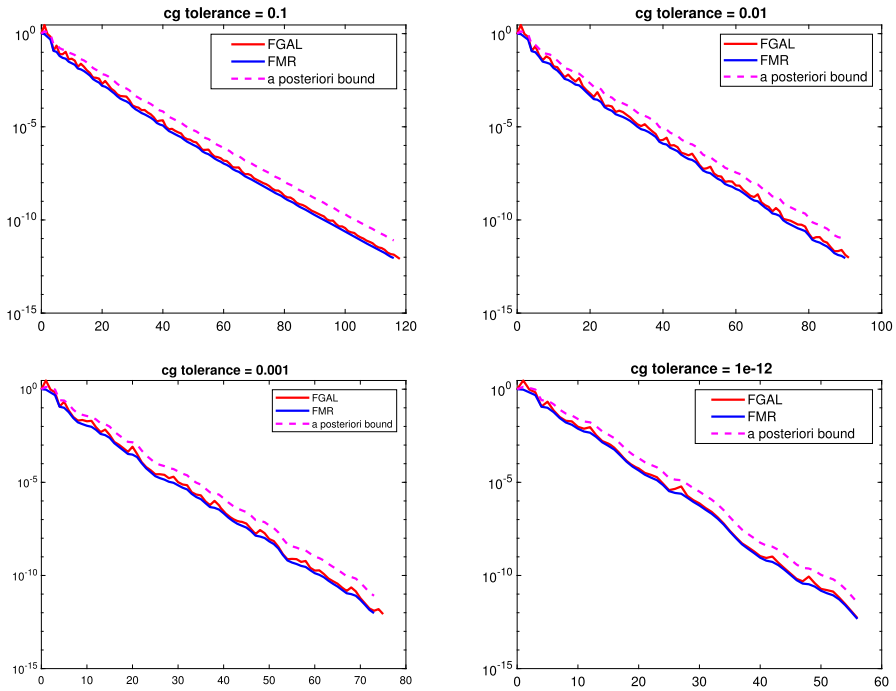
Fig. 2 Results for the unsteady channel domain Stokes problem, Example 2, $\tau/2 = 10^{-3}$, four choices for ε_{cg}

Example 2 This is the matrix treated in section 6.2 of [10]. It is based on a discretized Stokes problem obtained by using the Q_1-Q_1 finite element discretization of the unsteady channel domain problem in IFISS [25]. This example was treated for two different grid parameters in [10]. We only consider the larger of the two, i.e., grid parameter 9, resulting in a linear system of size $n = 789, 507$, and took the same stabilization matrix as in [10]. Finally, the step size in the midpoint Euler rule is set by $\tau/2 = 10^{-3}$.

The results are reported in Fig. 2. To solve the systems with H we use preconditioned CG with the modified no-fill ILU preconditioner. In the solves for H, every about 14 preconditioned CG iterations reduce the norm of the residual by another factor of 10^{-1}.

We observe that the dependence on the CG tolerance ε_{cg} is similar as in the previous example, with $\varepsilon_{cg} = 10^{-2}$ already requiring just about half times more the number of iterations than with "exact" solves ($\varepsilon_{cg} = 10^{-12}$). FGAL and FMR behave similarly with FGAL showing a slightly more oscillatory convergence behavior than before.

Two points are particularly interesting when comparing these results with those reported in [10]. The first is that the convergence history for FGAL is oscillatory in [10] with the residual norms decreasing from each iteration to the second-to-next, while increasing in-between. The plots in [10] report the 2-norms $\sqrt{r^*r}$ of the residuals r for the left preconditioned methods, whereas we perform right preconditioning and report approximate H^{-1}-norms, i.e., $\sqrt{r^*\tilde{r}}$ with $\tilde{r}$ the approximation for $H^{-1}r$ obtained

with preconditioned CG. These quantities decrease in a much smoother manner. The second point is that [10] uses the full Cholesky decomposition of H to obtain accurate inversions with H. The Cholesky factor of H has about $4 \cdot 10^8$ non-zero entries, whereas our modified no-fill incomplete Cholesky factor is 100 times "smaller" with $3.9 \cdot 10^6$ non-zero entries. This information can be used to roughly estimate the gains to be obtained with the flexible method and, e.g., $\epsilon_{cg} = 10^{-1}$: One (outer) iteration in the flexible method is about $100/14 \approx 7$ times cheaper than in the non-flexible method with exact solves, since it requires 14 preconditioned CG iterations with an incomplete Cholesky factor which has 100 times less non-zeros than the exact Cholesky factor. Even though this gain per iteration has to be halved because we need about twice as many (outer) iterations in the flexible method with $\epsilon_{cg} = 10^{-1}$, it is still substantial. In addition, the flexible method needs two orders of magnitude fewer memory.

The remaining two examples are linear systems that arise when using the implicit midpoint Euler rule to integrate an ODE resulting from a port-Hamiltonian modelling. These ODEs have the form

$$E\dot{u}(t) = (J - R)u(t) + f(t), \ u(0) = u_0, \tag{17}$$

where E is Hermitian and positive definite, J is anti-selfadjoint and R is Hermitian and positive semidefinite. With a step size of $\tau > 0$, the linear systems to solve at each time step are then of the form

$$Ax = b \ \text{ with } A = E + \frac{\tau}{2}(R - J),$$

and we have $H = E + \frac{\tau}{2}R$, positive definite, and $S = -\frac{\tau}{2}J$. Note that implicit integration methods are mandatory in this situation if one wants to preserve the dissipation of energy; see [16].

Example 3 We take the homonomically constrained spring mass system from [17], which is part of the port-Hamiltonian system benchmark collection.[2] We refer to [11] and [10] on how the matrices E, R and J arise from the descriptor system formulation of the system. We took the system with $n = 2 \cdot 10^6$ as it was also considered in [10] and $\tau/2 = 0.1$. The numerical results are given in Fig. 3. For all choices of ε_{cg}, the norms of the residuals of FGAL and FMR are so close that they are indistinguishable in the plots, which is why we report only the results for FMR. The bounds from Proposition 5 again match the convergence behavior quite exactly. Similar to the previous two examples, we see that with $\varepsilon_{cg} = 10^{-1}$ we need less than twice as many iterations than with "exact" solves ($\varepsilon_{cg} = 10^{-12}$), and this number of iterations is this time exceeded by just 1 for $\varepsilon_{cg} = 10^{-2}$. We need between 4 and 6 (unpreconditioned) CG iterations with H to reduce the initial residual by the factor $\varepsilon_{cg} = 10^{-1}$ in each step of ℓMR, and between 43 and 49 to reduce it by the factor $\varepsilon_{cg} = 10^{-12}$.

In the three plots for $\varepsilon_{cg} = 10^{-1}$, 10^{-2} and 10^{-12} we in addition plot the H-norms of the residuals in a variant of FMR which differs from FMR only by the fact that in the
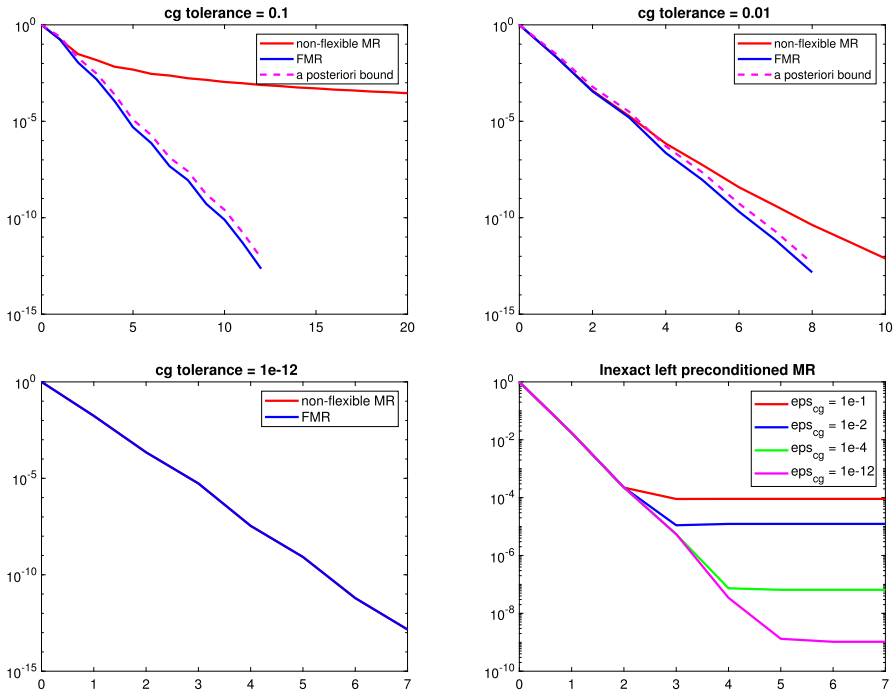

Fig. 3 Convergence plots for the spring mass system of dimension 10^6, Example 3, $\tau/2 = 10^{-1}$, three choices for ε_{cg}. Bottom right plot is for non-flexible ℓMR with inexact solves for H

underlying preconditioned Lanczos algorithm, Algorithm 2, we do not compute γ_k but rather set $\gamma_k = \beta_{k-1}$. This method is termed "non-flexible MR". The plots show that for the larger values of ε_{cg} this modification results in a significantly to dramatically delayed convergence.

In a similar spirit, the bottom right plot of Fig. 3 show the H^{-1}-norm of the residual for the non-flexible ℓMR method of Rapoport (left preconditioning), where we solve for H with CG at different accuracies. Unsurprisingly, albeit convincingly, the plot illustrates that this "inexact" version of ℓMR stagnates the sooner the less accurately we solve for H. A similar behavior can be observed for non-flexible inexact ℓGAL, for which we do not report results here. In contrast, the flexible methods reach high accuracies for the overall system even when ε_{cg} is large.

Example 4 We take the coupled electro-thermal DAE system from [2], from which we treat the thermal part after applying the decoupling described in [2]. This again yields an ODE of the form (17). Similar to the previous example, we consider the linear systems arising when using the implicit midpoint Euler method for time integration. The system has dimension $n = 7.2 \times 10^5$, and we take $\tau/2 = 10^{-2}$. Since the positive definite diagonal matrix E is highly ill-conditioned, we scaled the system from left and right with the square root of the inverse of E.

The results are reported in Fig. 4. To solve the systems with H we use preconditioned CG with a threshold dropping incomplete Cholesky preconditioner with drop tolerance
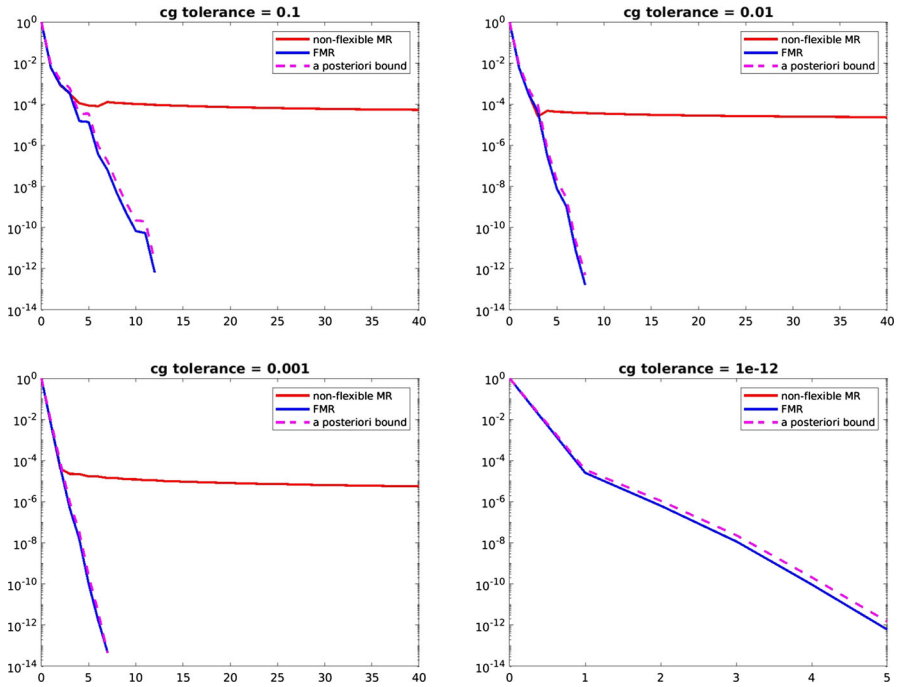
Fig. 4 Results for the coupled electro-thermal problem in Example 4 with $\tau/2 = 10^{-2}$, four choices for ε_{cg}

10^{-4}. This results in a moderately efficient preconditioner, since we now roughly need between 5 and 100 iterations to decrease the residual by one order of magnitude. Similarly to the previous example, FMR and FGAL perform very similarly and would be indistinguishable in the plots, which is why we only report the results for FMR. As before, we observe that FMR (and FGAL) require about twice the iterations with $\varepsilon_{cg} = 10^{-1}$ than with "exact" solves ($\varepsilon_{cg} = 10^{-12}$). Both flexible methods again achieve small final residuals even when large values of ε_{cg} are used.

For non-flexible MR, the iterations stagnate quite early, except for $\varepsilon_{cg} = 10^{-12}$, in which case the non-flexible method cannot be distinguished from FMR.

6 Conclusions

We presented new flexible versions of two known non-flexible Krylov subspace methods, based on right preconditioning with the Hermitian part H of the matrix which is assumed to be positive definite. The iterates of the first method are characterized by the Galerkin variational principle, the iterates of the second by a minimal residual property. Both methods rely on the H^{-1} inner product. They have the advantage of requiring only short recurrences, and the crucial feature that they allow for inexact solutions of systems with H in every iteration. We presented a theoretical analysis of the convergence of the flexible methods and a way to a posteriori compute fairly

precise approximations for error bounds on the H^{-1}-norm of the residuals. Numerical experiments show that the methods are particularly well suited for implicit time stepping methods for port-Hamiltonian ODEs and ODEs with a dissipative Hamiltonian. Future work may include the extension to DAEs where H is only positive semidefinite.

Data availability Data sharing not applicable to this article as no datasets were generated or analyzed during the current study.

Declarations

Conflict of interest The authors declare no competing interests.

References

1. Arioli, M., Liesen, J., Miedlar, A., Strakoš, Z.: Interplay between discretization and algebraic computation in adaptive numerical solution of elliptic PDE problems. GAMM-Mitt **36**(1), 102–129 (2013). https://doi.org/10.1002/gamm.201310006
2. Banagaaya, N., Feng, L., Schoenmaker, W., Meuris, P., Wieers, A., Gillon, R., Benner, P.: Model order reduction for nanoelectronics coupled problems with many inputs. In: Fanucci, L., Teich, J. (eds.) 2016 Design, Automation & Test in Europe Conference & Exhibition (DATE), pp. 313–318 (2016)
3. Concus, P., Golub, G.H.: A generalized conjugate gradient method for nonsymmetric systems of linear equations. In: Glowinski, R., Lions, J.L. (eds.) Computing Methods in Applied Sciences and Engineering (Second Internat. Sympos., Versailles, 1975), Part 1, Lect. Notes Econ. Math. Syst., vol. 134, Springer, pp. 56–65 (1976)
4. Eisenstat, S.C.: A note on the generalized conjugate gradient method. SAIM J. Numer. Anal. **20**, 358–362 (1983). https://doi.org/10.1137/0720024
5. Faber, V., Manteuffel, T.: Necessary and sufficient conditions for the existence of a conjugate gradient method. SIAM J. Numer. Anal. **21**(2), 352–362 (1984). https://doi.org/10.1137/0721026
6. Freund, R.: On conjugate gradient type methods and polynomial preconditioners for a class of complex non-Hermitian matrices. Numer. Math. **57**, 285–312 (1990). https://doi.org/10.1007/BF01386412
7. Freund, R.W., Nachtigal, N.M.: QMR: a quasi-minimal residual method for non-Hermitian linear systems. Numer. Math. **60**(3), 315–339 (1991). https://doi.org/10.1007/BF01385726
8. Fridman, V.M.: The method of minimal iterations with minimal errors for a system of linear algebraic equations with symmetric matrix. Ž Vyčisl Mat i Mat Fiz **2**, 341–342 (1962)
9. Golub, G.H., Ye, Q.: Inexact preconditioned conjugate gradient method with inner-outer iteration. SIAM J. Sci. Comput. **21**(4), 1305–1320 (1999). https://doi.org/10.1137/S1064827597323415

10. Güdücü, C., Liesen, J., Mehrmann, V., Szyld, D.B.: On non-Hermitian positive (semi)definite linear algebraic systems arising from dissipative Hamiltonian DAEs. SIAM J. Sci. Comput. **44**(4), A2871–A2894 (2022). https://doi.org/10.1137/21M1458594
11. Gugercin, S., Polyuga, R.V., Beattie, C., van der Schaft, A.: Structure-preserving tangential interpolation for model reduction of port-Hamiltonian systems. Autom. J. IFAC **48**(9), 1963–1974 (2012). https://doi.org/10.1016/j.automatica.2012.05.052
12. Hageman, L.A., Luk, F.T., Young, D.M.: On the equivalence of certain iterative acceleration methods. SIAM J. Numer. Anal. **17**, 852–873 (1980). https://doi.org/10.1137/0717071
13. Liesen, J.: When is the adjoint of a matrix a low degree rational function in the matrix? SIAM J. Matrix Anal. Appl. **29**(4), 1171–1180 (2007). https://doi.org/10.1137/060675538
14. Liesen, J., Strakoš, Z.: Krylov Subspace Methods. Numerical Mathematics and Scientific Computation—Principles and Analysis. Oxford University Press, Oxford (2013)
15. Málek J, Strakoš Z (2015) Preconditioning and the Conjugate Gradient Method in the Context of Solving PDEs, SIAM Spotlights, vol 1. Society for Industrial and Applied Mathematics (SIAM), Philadelphia
16. Mehrmann V, Morandin R (2019) Structure-preserving discretization for port-Hamiltonian descriptor systems. https://doi.org/10.48550/arXiv.1903.10451,
17. Mehrmann, V., Stykel, T.: Balanced truncation model reduction for large-scale systems in descriptor form. In: Benner, P., Mehrmann, V., Sorensen, D. (eds.) Dimension Reduction for Large Scale Systems, Lect. Notes Comput. Sci. Eng., vol. 45, Springer, Berlin, pp. 83–115 (2005). https://doi.org/10.1007/3-540-27909-1_3
18. Meurant, G., Papež, J., Tichý, P.: Accurate error estimation in CG. Numer. Algorithms **88**(3), 1337–1359 (2021). https://doi.org/10.1007/s11075-021-01078-w
19. Notay, Y.: Flexible conjugate gradients. SIAM J. Sci. Comput. **22**(4), 1444–1460 (2000). https://doi.org/10.1137/S1064827599362314
20. Paige, C.C., Saunders, M.A.: Solutions of sparse indefinite systems of linear equations. SIAM J. Numer. Anal. **12**(4), 617–629 (1975). https://doi.org/10.1137/0712047
21. Rapoport, D.: A Nonlinear Lanczos Algorithm and the Stationary Navier-Stokes Equation. ProQuest LLC, Ann Arbor (1978). Ph.D. thesis, New York University. http://gateway.proquest.com/openurl?url_ver=Z39.88-2004&rft_val_fmt=info:ofi/fmt:kev:mtx:dissertation&res_dat=xri:pqdiss&rft_dat=xri:pqdiss:7912313
22. Saad, Y.: A flexible inner-outer preconditioned GMRES algorithm. SIAM J. Sci. Comput. **14**(2), 461–469 (1993). https://doi.org/10.1137/0914028
23. Saad, Y.: Iterative Methods for Sparse Linear Systems, 2nd edn. Society for Industrial and Applied Mathematics, Philadelphia (2003). https://doi.org/10.1137/1.9780898718003
24. Sarkis, M., Szyld, D.B.: Optimal left and right additive Schwarz preconditioning for minimal residual methods with Euclidean and energy norms. Comput. Methods Appl. Mech. Eng. **196**, 1612–1621 (2007). https://doi.org/10.1016/j.cma.2006.03.027
25. Silvester, D., Elman, H.C., Ramage, A.: Incompressible flow and iterative solver software (IFISS), version 3.5 (2016). http://www.manchester.ac.uk/ifiss/
26. Simoncini, V., Szyld, D.B.: Recent computational developments in Krylov subspace methods for linear systems. Numer. Linear Algebra Appl. **14**, 1–59 (2007). https://doi.org/10.1002/nla.499
27. Szyld, D.B., Vogel, J.A.: FQMR: a flexible quasi-minimal residual method with inexact preconditioning. SIAM J. Sci. Comput. **23**(2), 363–380 (2001). https://doi.org/10.1137/S106482750037336X
28. Szyld, D.B., Widlund, O.: Variational analysis of some conjugate gradient methods. East-West J. Numer. Math. **21**, 51–74 (1993)
29. Trottenberg, U., Oosterlee, C.W., Schüller, A.: Multigrid. Academic Press Inc, San Diego (2001)
30. van der Vorst, H.A., Vuik, C.: GMRESR: a family of nested GMRES methods. Numer Linear Algebra Appl (1994). https://doi.org/10.1002/nla.1680010404
31. Widlund, O.: A Lanczos method for a class of nonsymmetric systems of linear equations. SIAM J. Numer. Anal. **15**(4), 801–812 (1978). https://doi.org/10.1137/0715053