Low-Rank Updates of Matrix Functions

Bernhard Beckermann, Daniel Kressner, Marcel Schweitzer

LOW-RANK UPDATES OF MATRIX FUNCTIONS

BERNHARD BECKERMANN*, DANIEL KRESSNER†, AND MARCEL SCHWEITZER‡

Abstract. We consider the task of updating a matrix function $f(A)$ when the matrix $A \in \mathbb{C}^{n\times n}$ is subject to a low-rank modification. In other words, we aim at approximating $f(A+D)-f(A)$ for a matrix D of rank $k \ll n$. The approach proposed in this paper attains efficiency by projecting onto tensorized Krylov subspaces produced by matrix-vector multiplications with A and A^*. We prove the approximations obtained from m steps of the proposed methods are exact if f is a polynomial of degree at most m and use this as a basis for proving a variety of convergence results, in particular for the matrix exponential and for Markov functions. We illustrate the performance of our method by considering various examples from network analysis, where our approach can be used to cheaply update centrality and communicability measures.

Key words. matrix function, low-rank update, Krylov subspace method, tensorized Krylov subspace, matrix exponential, Markov function, graph communicability

AMS subject classifications. 15A16, 65D30, 65F30, 65F60

1. Introduction.

This work is concerned with the problem of updating a matrix function $f(A)$ when $A \in \mathbb{C}^{n\times n}$ is subject to a low-rank modification. More specifically, given $D \in \mathbb{C}^{n\times n}$ of rank $r \ll n$ we aim at computing the difference

$$f(A+D)-f(A), \tag{1.1}$$

at a cost significantly smaller than the cost of computing $f(A+D)$ from scratch. Throughout this work, we assume that $f : \Omega \to \mathbb{C}$ is analytic on some domain Ω containing $\Lambda(A)$ and $\Lambda(A+D)$, the spectra of A and $A+D$.

The generality of (1.1) includes a wide scope of applications. We will particularly focus on measuring network properties, such as centrality and communicability, via applying the matrix exponential to the adjacency matrix of an undirected graph [18]. Removing/inserting individual edges or nodes in the graph corresponds to rank-2 updates of A. Being able to solve (1.1) efficiently therefore allows to quickly update these measures. In fact, this task is explicitly needed when measuring the betweenness of a node [18, Sec. 2.3].

For $f(z) = z^{-1}$, the well-known Sherman-Morrison-Woodbury formula implies that the difference (1.1) has rank r and can be directly obtained from applying $f(A) = A^{-1}$ to the low-rank factors of D. In particular, when D has rank 1 and can be written as $D = \boldsymbol{b}\boldsymbol{c}^*$ with vectors $\boldsymbol{b}$ and $\boldsymbol{c}$, we have

$$(A+\boldsymbol{b}\boldsymbol{c}^*)^{-1} - A^{-1} = -\frac{1}{1+\boldsymbol{c}^*A^{-1}\boldsymbol{b}} A^{-1}\boldsymbol{b}\boldsymbol{c}^*A^{-1}. \tag{1.2}$$

Unless A is a scalar multiple of the identity matrix [27, Theorem 1.35], there is no such simple relation between $f(A)$ and $f(A+\boldsymbol{b}\boldsymbol{c}^*)$ for a general analytic function f. In the case of a rational function $f \equiv r$ of degree δ, Bernstein and Van Loan [11] show

*Laboratoire Painlevé UMR 8524, UFR Mathématiques, Univ. Lille, 59655 Villeneuve d'Ascq, France. E-mail: bbecker@math.univ-lille1.fr. Supported in part by the Labex CEMPI (ANR-11-LABX-0007-01).

†MATHICSE-ANCHP, École Polytechnique Fédérale de Lausanne, Station 8, 1015 Lausanne, Switzerland. E-mail: daniel.kressner@epfl.ch

‡MATHICSE-ANCHP, École Polytechnique Fédérale de Lausanne, Station 8, 1015 Lausanne, Switzerland. E-mail: marcel.schweitzer@epfl.ch. The work of Marcel Schweitzer has been supported by the SNSF project *Low-rank updates of matrix functions and fast eigenvalue solvers.*

that $r(A + \boldsymbol{b}\boldsymbol{c}^*) - r(A)$ has rank at most δ and provide an explicit formula for the rank-δ correction. This construction is based on explicit Krylov bases and potentially prone to numerical instability for larger degrees.

The Cauchy integral representation

$$f(A) = \frac{1}{2\pi \mathrm{i}} \int_\Gamma f(z)(zI - A)^{-1} \mathrm{d}z \tag{1.3}$$

provides a link between matrix functions and (shifted) inverses. In [43], a combination of numerical quadrature with (1.2) has been explored for approximating $\exp(A + \boldsymbol{b}\boldsymbol{c}^*) - \exp(A)$. However, such an approach suffers from a number of drawbacks. Most importantly, the choice of a contour Γ that is equally good for A and $A + \boldsymbol{b}\boldsymbol{c}^*$ is highly nontrivial, in particular in the non-Hermitian case. Also, the evaluation of the approximation amounts to the solution of a differently shifted linear system for each quadrature point. Although not used in our algorithms, integral representations will play a role in their convergence analysis, see Section 5.

The approach proposed in this work avoids the need for choosing a contour or an explicit rational approximation of f. Inspired by existing Krylov subspace methods for matrix equations [41] and linear systems with tensor product structure [31], we make use of tensorized subspaces. More specifically, given orthonormal bases U_m, V_m for m-dimensional Krylov subspaces involving the matrices A, A^* and the starting vectors $\boldsymbol{b}, \boldsymbol{c}$, we construct an approximation of the form

$$f(A + \boldsymbol{b}\boldsymbol{c}^*) - f(A) \approx U_m X_m(f) V_m^*,$$

with a well-chosen small matrix $X_m(f) \in \mathbb{C}^{m \times m}$. In turn, when $m \ll n$, the computational cost is dominated by m matrix-vector products with A and A^*. As we will demonstrate for a variety of examples, this dramatically reduces the computational effort compared to computing $f(A + \boldsymbol{b}\boldsymbol{c}^*)$ from scratch, even when only selected quantities – such as the diagonal or trace of $f(A + \boldsymbol{b}\boldsymbol{c}^*)$ – are required.

The rest of this paper is organized as follows. In section 2, we introduce Krylov subspace algorithms for approximating the update (1.1) when D is a matrix of rank one, both for the Hermitian and the non-Hermitian case, and discuss the extension to higher rank. In section 3, we prove that the approximation obtained from m steps of our Krylov subspace algorithm is exact when f is a polynomial of degree at most m. This result forms the basis of the convergence analysis presented in sections 4 and 5. Precisely, the convergence of our Krylov approximations is linked to certain polynomial approximation problems in section 4, while section 5 contains results that are based on exploiting an integral representation of $f(A)$, e.g., for general analytic functions and for Markov functions. Applications for our methods are described in Section 6, with special focus on up-/down-dating of communicability measures in network analysis, and the efficiency of our approach is illustrated by numerical experiments from this area. Concluding remarks are given in Section 7.

We use $\|\cdot\|$ to denote the Euclidean norm of a vector or the induced spectral norm of a matrix.

2. Krylov projection algorithms. In the following, we assume that $f(A)$ or at least the part relevant to the application (e.g., the diagonal of $f(A)$) have already been computed. For the moment, we continue to assume a rank-1 modification, that is, we consider the approximation of $f(A + \boldsymbol{b}\boldsymbol{c}^*) - f(A)$ for vectors $\boldsymbol{b}, \boldsymbol{c}$. In Remark 2.3 below, we will comment on the extension to higher ranks.

2.1. Hermitian rank-1 updates. Because it is conceptually simpler, we first discuss the Hermitian case: $A = A^*$ and $\boldsymbol{b} = \boldsymbol{c}$.

The first step of our algorithm consists of constructing an orthonormal basis for a Krylov subspace of A with starting vector $\boldsymbol{b}$. Given $m \leq n$, such a Krylov subspace takes the form

$$\mathcal{U}_m := \mathcal{K}_m(A, \boldsymbol{b}) = \operatorname{span}\{\boldsymbol{b}, A\boldsymbol{b}, A^2\boldsymbol{b}, \ldots, A^{m-1}\boldsymbol{b}\}.$$

For simplicity, we suppose that $\mathcal{U}_m$ has dimension m, which is generically satisfied. Applying m steps of the Lanczos method [32] (possibly with reorthogonalization) results in the Lanczos relation

$$AU_m = U_m G_m + \beta_{m+1}\boldsymbol{u}_{m+1}\boldsymbol{e}_m^*, \tag{2.1}$$

where the columns of $U_m \in \mathbb{C}^{n\times m}$ form an orthonormal basis of $\mathcal{U}_m$, the matrix $G_m = U_m^* A U_m \in \mathbb{C}^{m\times m}$ is tridiagonal, and $\boldsymbol{e}_m$ denotes the mth unit vector of length m. The first column of U_m is a scalar multiple of $\boldsymbol{b}$ and, without loss of generality, we may assume that $U_m^*\boldsymbol{b} = \|\boldsymbol{b}\|\boldsymbol{e}_1$ with the first unit vector $\boldsymbol{e}_1 \in \mathbb{C}^m$.

The second step of our algorithm then chooses an approximation in the tensorized subspace $\mathcal{U}_m \otimes \mathcal{U}_m$, that is,

$$f(A + \boldsymbol{b}\boldsymbol{b}^*) - f(A) \approx U_m X_m(f) U_m^* \tag{2.2}$$

for some matrix $X_m(f) \in \mathbb{C}^{m\times m}$. In the spirit of Krylov subspace methods for matrix equations [41], we choose $X_m(f)$ as the solution of the compressed problem:

$$X_m(f) = f\big(U_m^*(A + \boldsymbol{b}\boldsymbol{b}^*)U_m\big) - f\big(U_m^* A U_m\big) = f\big(G_m + \|\boldsymbol{b}\|^2 \boldsymbol{e}_1\boldsymbol{e}_1^*\big) - f(G_m), \tag{2.3}$$

where we assume that f is defined on the spectra of G_m and $G_m + \|\boldsymbol{b}\|^2\boldsymbol{e}_1\boldsymbol{e}_1^*$. Below, in Theorem 3.2, we will see that this choice of $X_m(f)$ leads to an approximation that is exact when f is a polynomial of degree at most m.

The resulting method is summarized in Algorithm 1. The tridiagonal matrix G_m from (2.1) is built from the orthogonalization coefficients α_j, β_j as

$$G_m = \begin{bmatrix} \alpha_1 & \beta_2 & & & \\ \beta_2 & \alpha_2 & \beta_3 & & \\ & \ddots & \ddots & \ddots & \\ & & \beta_{m-1} & \alpha_{m-1} & \beta_m \\ & & & \beta_m & \alpha_m \end{bmatrix}.$$

A trivial modification of this algorithm can be used to approximate $f(A - \boldsymbol{b}\boldsymbol{b}^*) - f(A)$.

Algorithm 1 represents the most basic form of the *Lanczos process*; in particular, it does not employ any reorthogonalization. It is well known that such short recurrences may suffer from severe loss of orthogonality in the presence of round-off errors, so that it can be advisable to use reorthogonalization strategies [37, 40]. The most straightforward to retain numerical orthogonality amongst the basis vectors is to store all basis vectors and perform *full reorthogonalization* in each step of the method.

An alternative is to perform no reorthogonalization at all. In contrast to, e.g., the CG method for linear systems, there are no short recurrences for the iterates $U_m X_m(f) U_m^*$ available. In turn, this requires to use a *two-pass* strategy for forming

Algorithm 1 Krylov subspace approximation of $f(A+\boldsymbol{b}\boldsymbol{b}^*)-f(A)$ for Hermitian A

$\boldsymbol{u}_0 = \boldsymbol{0}$.
$\boldsymbol{u}_1 = (1/\|\boldsymbol{b}\|)\boldsymbol{b}$.
$\beta_1 = 0$.
for $j = 1, \ldots, m$ **do**
    $\boldsymbol{w}_j = A\boldsymbol{u}_j - \beta_j \boldsymbol{u}_{j-1}$.
    $\alpha_j = \boldsymbol{u}_j^* \boldsymbol{w}_j$.
    $\boldsymbol{w}_j = \boldsymbol{w}_j - \alpha_j \boldsymbol{u}_j$.
    $\beta_{j+1} = \|\boldsymbol{w}_j\|$.
    $\boldsymbol{u}_{j+1} = (1/\beta_{j+1})\boldsymbol{w}_j$.
end for
Compute matrix function $X_m(f) = f\big(G_m + \|\boldsymbol{b}\|^2 \boldsymbol{e}_1 \boldsymbol{e}_1^*\big) - f(G_m)$
Return $U_m X_m(f) U_m^*$.

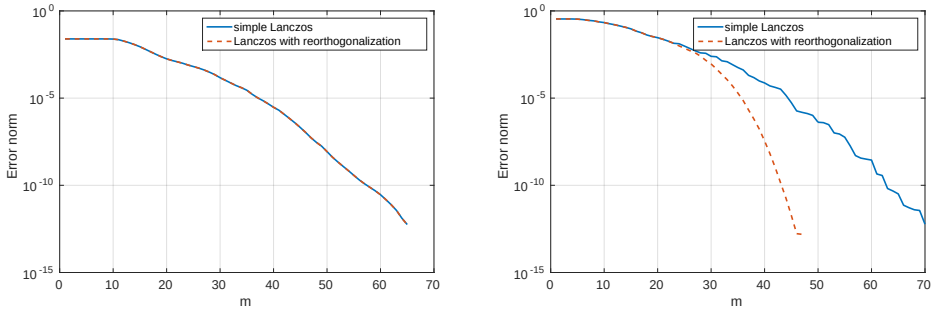

FIG. 2.1. *Convergence curves of simple Lanczos and Lanczos with full reorthogonalization for diagonal matrices with equidistantly spaced (left) and logarithmically spaced (right) eigenvalues in* $[10^{-3}, 10^3]$.

$U_m X_m(f) U_m^*$ if one wants to avoid storing the full basis U_m. In a first pass, the tridiagonal matrix G_m is assembled (while storing only three basis vectors at a time and discarding the older ones) and $X_m(f)$ is formed. In a second pass, the basis vectors are computed anew (again only storing three at a time), and the parts of interest of $U_m X_m(f) U_m^*$ (e.g., its diagonal) are gradually computed. This approach doubles the number of matrix-vector products but reduces the storage requirement from $\mathcal{O}(mn)$ to $\mathcal{O}(m^2 + n)$.

EXAMPLE 2.1. We illustrate the (possible) impact of reorthogonalization on convergence. Consider the diagonal matrices $A_1, A_2 \in \mathbb{C}^{100\times 100}$, where the eigenvalues of A_1 are equidistantly spaced in the interval $[10^{-3}, 10^3]$ and the eigenvalues of A_2 are logarithmically spaced in $[10^{-3}, 10^3]$. We approximate $f(A_i + \boldsymbol{b}\boldsymbol{b}^*) - f(A_i), i = 1, 2$ where $f(z) = \exp(-z)$ and $\boldsymbol{b}$ is a random vector of unit norm using Lanczos without reorthogonalization and with full reorthogonalization, as explained above. The resulting convergence curves are shown in Figure 2.1. For A_1, both methods are observed to behave identically, which is well explained by the fact that the eigenvalue distribution of A_1 does not favor the convergence of Ritz values, which has a close link to loss of orthogonality [35, 37]. For A_2, the eigenvalue distribution favors the convergence of Ritz values to larger eigenvalues and, in turn, convergence degrades after some time when using no reorthogonalization. Still, the method eventually converges to the same accuracy, it just needs more iterations. This phenomenon is

also known from the conjugate gradient method in finite precision arithmetic [35] and when approximating $f(A)\boldsymbol{b}$ [15]. To avoid that orthogonality issues distort our findings, we use the Lanczos method with full reorthogonalization for all numerical experiments in the rest of this paper. ◇

The computational cost of m steps of Algorithm 1 is as follows:

- m matrix-vector products with A ($2m$, when the two-pass approach is used), requiring $\mathcal{O}(m \cdot \mathrm{nnz}(A))$ operations for a sparse matrix A with $\mathrm{nnz}(A)$ nonzeros;
- $\mathcal{O}(m^2 n)$ operations for the orthogonalization procedure when using full reorthogonalization (and only $\mathcal{O}(mn)$ when the two-pass approach is used, because in each step orthogonalization against only two previous basis vectors is performed)
- the computation of $X_m(f)$, which requires the evaluation of two functions of $m \times m$ matrices, which, depending on the function f, typically needs at most $\mathcal{O}(m^3)$ operations.

One should avoid forming $U_m X_m(f) U_m^*$ explicitly and we expect that this is actually not needed in most applications involving a large sparse matrix A. For example, the computation of communicability measures discussed in Section 6.1 only requires the diagonal entries of $U_m X_m(f) U_m^*$, which can be computed directly from U_m, $X_m(f)$ with $\mathcal{O}(m^2 n)$ operations.

2.2. Non-Hermitian rank-1 updates. In the general non-Hermitian case, we construct orthonormal bases for the two (polynomial) Krylov subspaces $\mathcal{U}_m := \mathcal{K}_m(A, \boldsymbol{b})$ and $\mathcal{V}_m := \mathcal{K}_m(A^*, \boldsymbol{c})$. Applying the Arnoldi method with reorthogonalization results in the Arnoldi relations (2.1) and, additionally,

$$A^* V_m = V_m H_m + h_{m+1,m} \boldsymbol{v}_{m+1} \boldsymbol{e}_m^*, \tag{2.4}$$

where the columns of $V_m \in \mathbb{C}^{n \times m}$ form an orthonormal basis of $\mathcal{V}_m$. Note that both $G_m = U_m^* A U_m$ and $H_m = V_m^* A^* V_m$ are now $m \times m$ upper Hessenberg matrices. The approximation is chosen in the tensorized subspace $\mathcal{U}_m \otimes \mathcal{V}_m$, that is,

$$f(A + \boldsymbol{b}\boldsymbol{c}^*) - f(A) \approx U_m X_m(f) V_m^* \tag{2.5}$$

for some matrix $X_m(f) \in \mathbb{C}^{m \times m}$. Concerning the choice of $X_m(f)$, it turns out that the non-Hermitian analogue of (2.3) does not have favorable theoretical properties. For example, the polynomial exactness property mentioned above for (2.3) and proven in section 3 does not hold for such a choice. We will use a different choice, motivated by the following simple result.

LEMMA 2.2. *Let $A \in \mathbb{C}^{n \times n}$, and $\boldsymbol{b}, \boldsymbol{c} \in \mathbb{C}^n$. Define the block matrix*

$$\mathcal{A} := \begin{bmatrix} A & \boldsymbol{b}\boldsymbol{c}^* \\ 0 & A + \boldsymbol{b}\boldsymbol{c}^* \end{bmatrix}. \tag{2.6}$$

Then

$$f(\mathcal{A}) = \begin{bmatrix} f(A) & f(A + \boldsymbol{b}\boldsymbol{c}^*) - f(A) \\ 0 & f(A + \boldsymbol{b}\boldsymbol{c}^*) \end{bmatrix}.$$

Proof. Letting Γ denote a contour that encloses both $\Lambda(A)$ and $\Lambda(A+\boldsymbol{b}\boldsymbol{c}^*)$, the Cauchy integral formula (1.3) applied to $f(\mathcal{A})$ yields

$$\begin{aligned} f(\mathcal{A}) &= \frac{1}{2\pi i}\int_\Gamma f(z)(zI-\mathcal{A})^{-1}\,\mathrm{d}z \\ &= \frac{1}{2\pi i}\int_\Gamma f(z)\begin{bmatrix}(zI-A)^{-1} & -(zI-A)^{-1}\boldsymbol{b}\boldsymbol{c}^*(zI-A-\boldsymbol{b}\boldsymbol{c}^*)^{-1}\\ 0 & (zI-A-\boldsymbol{b}\boldsymbol{c}^*)^{-1}\end{bmatrix}\mathrm{d}z \\ &= \begin{bmatrix} f(A) & -\frac{1}{2\pi i}\int_\Gamma f(z)(zI-A)^{-1}\boldsymbol{b}\boldsymbol{c}^*(zI-A-\boldsymbol{b}\boldsymbol{c}^*)^{-1}\,\mathrm{d}z \\ 0 & f(A+\boldsymbol{b}\boldsymbol{c}^*)\end{bmatrix}. \end{aligned} \tag{2.7}$$

On the other hand, combining the Cauchy integral formula for $f(A+\boldsymbol{b}\boldsymbol{c}^*)$ and $f(A)$ with the second resolvent identity yields

$$\begin{aligned} f(A+\boldsymbol{b}\boldsymbol{c}^*)-f(A) &= \frac{1}{2\pi i}\int_\Gamma f(z)\big[(zI-A-\boldsymbol{b}\boldsymbol{c}^*)^{-1}-(zI-A)^{-1}\big]\,\mathrm{d}z \\ &= -\frac{1}{2\pi i}\int_\Gamma f(z)(zI-A-\boldsymbol{b}\boldsymbol{c}^*)^{-1}\boldsymbol{b}\boldsymbol{c}^*(zI-A)^{-1}\,\mathrm{d}z \\ &= -\frac{1}{2\pi i}\int_\Gamma f(z)(zI-A)^{-1}\boldsymbol{b}\boldsymbol{c}^*(zI-A-\boldsymbol{b}\boldsymbol{c}^*)^{-1}\,\mathrm{d}z, \end{aligned} \tag{2.8}$$

which matches the (1,2) block of (2.7) and thus completes the proof. ∎

The result of Lemma 2.2 shows that the desired update is contained in the (1,2) block of $f(\mathcal{A})$. This motivates us to choose $X_m(f)$ as the (1,2) block of f applied to the compression of $\mathcal{A}$ onto $\mathcal{U}_m \oplus \mathcal{V}_m$:

$$\begin{aligned} &f\left(\begin{bmatrix}U_m & 0\\ 0 & V_m\end{bmatrix}^*\begin{bmatrix}A & \boldsymbol{b}\boldsymbol{c}^*\\ 0 & A+\boldsymbol{b}\boldsymbol{c}^*\end{bmatrix}\begin{bmatrix}U_m & 0\\ 0 & V_m\end{bmatrix}\right) \\ &= f\left(\begin{bmatrix}G_m & \|\boldsymbol{b}\|\|\boldsymbol{c}\|\boldsymbol{e}_1\boldsymbol{e}_1^*\\ 0 & H_m^*+\|\boldsymbol{c}\|V_m^*\boldsymbol{b}\boldsymbol{e}_1^*\end{bmatrix}\right) = \begin{bmatrix}f(G_m) & X_m(f)\\ 0 & f(H_m^*+\|\boldsymbol{c}\|V_m^*\boldsymbol{b}\boldsymbol{e}_1^*)\end{bmatrix}, \end{aligned} \tag{2.9}$$

where we again assume that f is defined on the spectra of G_m and $H_m^*+\|\boldsymbol{c}\|V_m^*\boldsymbol{b}\boldsymbol{e}_1^*$. Using Lemma 2.2, it is straightforward to see that this choice of $X_m(f)$ coincides in the Hermitian case with the one from section 2.1.

Algorithm 2 summarizes the proposed procedure, where we omit the algorithmic details for the Arnoldi method for the sake of brevity; see, e.g., [21].

Algorithm 2 Krylov subspace approximation of $f(A+\boldsymbol{b}\boldsymbol{c}^*)-f(A)$

1: Perform m steps of the Arnoldi method to compute an orthonormal basis U_m of $\mathcal{K}_m(A,\boldsymbol{b})$ and $G_m = U_m^*AU_m$.
2: Perform m steps of the Arnoldi method to compute an orthonormal basis V_m of $\mathcal{K}_m(A^*,\boldsymbol{c})$ and $H_m = V_m^*A^*V_m$.
3: Compute matrix function $F_m = f\left(\begin{bmatrix}G_m & \|\boldsymbol{b}\|\|\boldsymbol{c}\|\boldsymbol{e}_1\boldsymbol{e}_1^*\\ 0 & H_m^*+\|\boldsymbol{c}\|V_m^*\boldsymbol{b}\boldsymbol{e}_1^*\end{bmatrix}\right)$.
4: Set $X_m(f) = F_m(1:m, m+1:2m)$.
5: Return $U_mX_m(f)V_m^*$.

The computational effort of Algorithm 2 compares to that of Algorithm 1 (with full reorthogonalization) as follows. In contrast to the Hermitian case, we now need to build *two* Krylov spaces instead of one, meaning that the number of matrix vector

products necessary for m steps of the method increases from m to $2m$. The orthogonalization cost is the same as in the Hermitian case (with full reorthogonalization). As the number of operations needed for approximating a (dense) matrix function often grows cubically in the matrix size it is typically about four times as expensive to compute one matrix function of size $2m \times 2m$ than computing two matrix functions of size $m \times m$. The fact that the matrix for which we need to evaluate f in Algorithm 2 is non-Hermitian and not tridiagonal may increase the cost further. However, as long as $m \ll n$, the main cost of Algorithm 2 consists of performing matrix-vector products, so that we can expect it to take roughly twice the computation time of Algorithm 1 for a problem of the same size and structure.

REMARK 2.3. There are two different ways of extending our approach from a rank-one modification to a general rank-k modification $f(A+D)$ with $D = BC^*$ and $B, C \in \mathbb{C}^{n\times k}$:

1. By letting $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_k$ and $\boldsymbol{c}_1, \ldots, \boldsymbol{c}_k$ denote the columns of B and C, respectively, we can write D as a sum of k rank-one matrices, $D = \sum_{i=1}^{k} \boldsymbol{b}_i \boldsymbol{c}_i^*$, e.g., by computing a singular value decomposition of D and then apply k times Algorithm 2 in order to subsequently incorporate each of the k rank-one modifications. Note that the ith step of this procedure requires working with the Krylov subspaces $\mathcal{K}_m(A+\sum_{j=1}^{i-1} \boldsymbol{b}_j \boldsymbol{c}_j^*, \boldsymbol{b}_i)$ and $\mathcal{K}_m(A^*+\sum_{j=1}^{i-1} \boldsymbol{c}_j \boldsymbol{b}_j^*, \boldsymbol{c}_i)$ which in general do *not* coincide with $\mathcal{K}_m(A, \boldsymbol{b}_i)$ and $\mathcal{K}_m(A^*, \boldsymbol{c}_i)$.

Some care is required in order to make use of Algorithm 1 for a Hermitian rank-k matrix D. After a preprocessing step (see, e.g., [6, Section 2.3]) one can write $D = \sum_{i=1}^{\tilde{k}} \boldsymbol{b}_i \boldsymbol{b}_i^* - \sum_{i=\tilde{k}+1}^{k} \boldsymbol{b}_i \boldsymbol{b}_i^*$. First, Algorithm 1 is applied to incorporate the first $\tilde{k}$ terms followed by a slight modification of this algorithm to incorporate the last $k - \tilde{k}$ terms

2. *Block Krylov subspaces* [24,36,42] offer a conceptually different way of dealing with a rank-k modification. Instead of the Arnoldi method, a block Arnoldi method is used for computing orthonormal bases of the block Krylov subspaces $\mathcal{K}_m(A, B) = \text{range}[B, AB, \ldots, A^{m-1}B]$ and $\mathcal{K}_m(A^*, C)$. The approximation of $f(A + D) - f(A)$ is computed by projecting onto the tensorized block Krylov subspace $\mathcal{K}_m(A, B) \otimes \mathcal{K}_m(A^*, C)$, leading to a straightforward extension of Algorithm 2. Block Krylov subspace methods are more complicated to implement, see, e.g., [24] for some of the issues one has to take into account, but they offer (at least) one major advantage. Even though the product of A with an $n \times k$ matrix requires the same number of operations as k individual matrix vector products, it often performs much faster on a computer, benefitting from a more "cache-friendly" memory access pattern, see, e.g., [3]. Again some care is required in the symmetric case, to derive a block variant of Algorithm 1.

2.3. Stopping criterion. A simple stopping criterion for Algorithm 1 or Algorithm 2 is based on the error estimate

$$\|f(A + \boldsymbol{b}\boldsymbol{c}^*) - U_m X_m(f) V_m^*\| \approx \|U_{m+d} X_{m+d}(f) V_{m+d}^* - U_m X_m(f) V_m^*\| \qquad (2.10)$$

for some small integer $d \geq 1$. This error estimate is similar in spirit to error estimates for linear systems proposed in [22, 34]. Evaluating the right-hand side of (2.10) only requires forming the coefficient matrices $X_m(f), X_{m+d}(f)$ defined in (2.2) or (2.5),

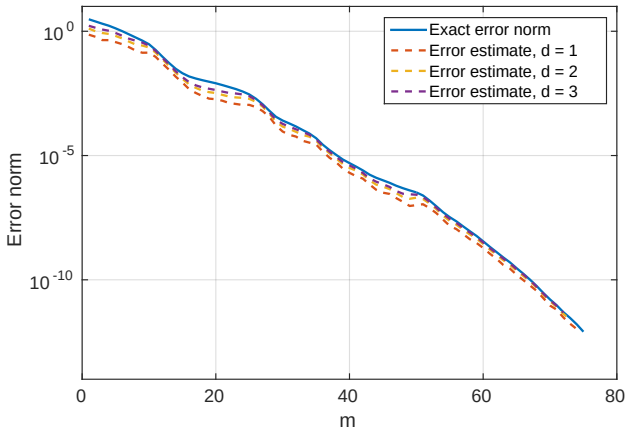

FIG. 2.2. *Comparison of exact error and difference-based error estimates for the update of the inverse square root described in Example 2.4.*

because

$$\begin{aligned}
&\|U_{m+d}X_{m+d}(f)V_{m+d}^* - U_mX_m(f)V_m^*\| \\
&= \left\| U_{m+d}X_{m+d}(f)V_{m+d}^* - U_{m+d}\begin{bmatrix} X_m(f) & 0 \\ 0 & 0 \end{bmatrix} V_{m+d}^* \right\| \\
&= \left\| X_{m+d}(f) - \begin{bmatrix} X_m(f) & 0 \\ 0 & 0 \end{bmatrix} \right\|.
\end{aligned}$$

We illustrate the behavior of the error estimator (2.10) by means of a simple numerical experiment.

EXAMPLE 2.4. We choose $A \in \mathbb{R}^{400\times 400}$ as the standard finite difference discretization of the two-dimensional Laplace operator on the unit grid, and $\boldsymbol{b}, \boldsymbol{c}$ as vectors with random, normal distributed entries. We use our proposed Krylov subspace algorithm to approximate $(A + \boldsymbol{b}\boldsymbol{c}^*)^{-1/2} - A^{-1/2}$ and show the convergence history as well as the proposed error estimator for $d = 1, 2, 3$ in Figure 2.2. We observe that the difference-based error estimators follow the norm of the exact error very closely and thus give very accurate results already for such small values of d. $\diamond$

We remark that the error estimator (2.10) is clearly heuristic and can be overly optimistic, especially in situations where the iteration (almost) stagnates, although it works very well in most practical situations. Of course, when already $m + d$ iterations of the method have been performed, one will typically use the approximation $U_{m+d}X_{m+d}V_{m+d}^*$ from the $(m+d)$th step instead of the approximation from the mth step (for which the error is estimated), as it can be expected to be a more accurate approximation. In fact, in Example 2.4, for $d = 3$ the error of the iterate X_{m+d} lies *below* the error estimate (2.10) for all but four iterations.

3. Exactness properties of Krylov subspace approximations for the update. In this section, we show that the approximation (2.5) is exact when f is a polynomial of degree at most m. This serves two purposes: On the one hand, this justifies the particular choice of approximation we made. On the other hand, this provides the fundamental basis for performing our convergence analyses in section 4.

As a tool for proving polynomial exactness, we utilize the following matrix identity, which does not seem to be widely known.

PROPOSITION 3.1. *Let $M, N \in \mathbb{C}^{n\times n}$. Then, for any $j \in \mathbb{N}$,*

$$M^j - N^j = \sum\nolimits_{k=0}^{j-1} N^{j-1-k}(M-N)M^k.$$

Proof. The proof is by induction on j. For $j = 0$, the assertion trivially holds. Suppose now that the assertion holds for $j-1$:

$$M^{j-1} - N^{j-1} = \sum\nolimits_{k=0}^{j-2} N^{j-2-k}(M-N)M^k. \tag{3.1}$$

Multiplying both sides of (3.1) by N from the left and adding M^j, we find

$$M^j + NM^{j-1} - N^j = M^j + \sum\nolimits_{k=0}^{j-2} N^{j-1-k}(M-N)M^k.$$

Finally, subtracting $M^{j-1}N$ on both sides and noting that $M^j - NM^{j-1} = N^0(M-N)M^{j-1}$ completes the proof. ∎

THEOREM 3.2. *Let $A \in \mathbb{C}^{n\times n}$, $\boldsymbol{b}, \boldsymbol{c} \in \mathbb{C}^n$. Then the Krylov subspace approximation returned by Algorithm 2 is exact for all $p \in \Pi_m$, where Π_m denotes the space of all polynomials of degree at most m, i.e.,*

$$p(A + \boldsymbol{b}\boldsymbol{c}^*) - p(A) = U_m X_m(p) V_m^*.$$

Proof. By linearity, it suffices to show that the result holds for every monomial $p_j(z) = z^j$, $j = 0, \dots, m$. We recall that $X_m(p_j)$ is the $(1,2)$-block of the matrix

$$\begin{bmatrix} G_m & \|\boldsymbol{b}\|\|\boldsymbol{c}\|\boldsymbol{e}_1\boldsymbol{e}_1^* \\ 0 & (H_m^* + \|\boldsymbol{c}\|V_m^*\boldsymbol{b}\boldsymbol{e}_1^*) \end{bmatrix}^j.$$

In particular, $X_m(p_1) = \|\boldsymbol{b}\|\|\boldsymbol{c}\|\boldsymbol{e}_1\boldsymbol{e}_1^*$. By (2.9), the result of the theorem is shown if we can establish

$$(A + \boldsymbol{b}\boldsymbol{c}^*)^j - A^j = U_m X_m(p_j) V_m^*. \tag{3.2}$$

Considering

$$\begin{bmatrix} G_m & \|\boldsymbol{b}\|\|\boldsymbol{c}\|\boldsymbol{e}_1\boldsymbol{e}_1^* \\ 0 & (H_m^* + \|\boldsymbol{c}\|V_m^*\boldsymbol{b}\boldsymbol{e}_1^*) \end{bmatrix}^j = \begin{bmatrix} G_m^{j-1} & X_m(p_{j-1}) \\ 0 & (H_m^* + \|\boldsymbol{c}\|V_m^*\boldsymbol{b}\boldsymbol{e}_1^*)^{j-1} \end{bmatrix} \begin{bmatrix} G_m & X_m(p_1) \\ 0 & (H_m^* + \|\boldsymbol{c}\|V_m^*\boldsymbol{b}\boldsymbol{e}_1^*) \end{bmatrix},$$

we find the relation

$$X_m(p_j) = G_m^{j-1} X_m(p_1) + X_m(p_{j-1})(H_m^* + \|\boldsymbol{c}\|V_m^*\boldsymbol{b}\boldsymbol{e}_1^*).$$

Resolving this recursion gives

$$X_m(p_j) = \sum\nolimits_{k=0}^{j-1} G_m^{j-1-k}(\|\boldsymbol{b}\|\|\boldsymbol{c}\|\boldsymbol{e}_1\boldsymbol{e}_1^*)(H_m^* + \|\boldsymbol{c}\|V_m^*\boldsymbol{b}\boldsymbol{e}_1^*)^k. \tag{3.3}$$

Recalling that G_m is the compression of A onto $\mathcal{K}_m(A, \boldsymbol{b})$, a well-known polynomial exactness property (see, e.g., [17, 39]) of Krylov subspace approximations yields

$$\|\boldsymbol{b}\| U_m G_m^\ell \boldsymbol{e}_1 = A^\ell \boldsymbol{b} \text{ for all } \ell = 0, \dots, m-1. \tag{3.4}$$

Similarly, by noting that $H_m + \|\boldsymbol{c}\|\boldsymbol{e}_1\boldsymbol{b}V_m^*$ is the compression of $(A + \boldsymbol{b}\boldsymbol{c}^*)^*$ onto $\mathcal{K}_m(A^*, \boldsymbol{c}) = \mathcal{K}_m((A + \boldsymbol{b}\boldsymbol{c}^*)^*, \boldsymbol{c})$, we obtain

$$\|\boldsymbol{c}\|\boldsymbol{e}_1^*(H_m^* + \|\boldsymbol{c}\|V_m^*\boldsymbol{b}\boldsymbol{e}_1^*)^\ell V_m^* = \boldsymbol{c}^*(A + \boldsymbol{b}\boldsymbol{c}^*)^\ell \text{ for all } \ell = 0, \ldots, m-1. \tag{3.5}$$

Multiplying (3.3) by U_m from the left and by V_m^* from the right then gives

$$U_m X_m(p_j) V_m^* = \sum\nolimits_{k=0}^{j-1} U_m G_m^{j-1-k}(\|\boldsymbol{b}\|\|\boldsymbol{c}\|\boldsymbol{e}_1\boldsymbol{e}_1^*)(H_m^* + \|\boldsymbol{c}\|V_m^*\boldsymbol{b}\boldsymbol{e}_1^*)^k V_m^*$$

which by (3.4)–(3.5) gives for $j \le m$ the relation

$$U_m X_m(p_j) V_m^* = \sum\nolimits_{k=0}^{j-1} A^{j-1-k}\boldsymbol{b}\boldsymbol{c}^*(A + \boldsymbol{b}\boldsymbol{c}^*)^k = (A + \boldsymbol{b}\boldsymbol{c}^*)^j - A^j,$$

where we used Proposition 3.1 in the second equality. This establishes (3.2) and thus completes the proof. ∎

4. Convergence analysis based on polynomial approximation problems. In this section, we give bounds for the error of the approximations (2.2) and (2.5). The results are based on the polynomial exactness property from Theorem 3.2 and connect the approximation quality of (2.2) and (2.5) to certain polynomial approximation problems.

The convergence results in this section rely on a theorem by Crouzeix [12,13], for which we first recall some basic concepts. The *field of values* (or *numerical range*) $A \in \mathbb{C}^{n\times n}$ is defined as the set

$$W(A) := \{\boldsymbol{x}^*A\boldsymbol{x} : \|\boldsymbol{x}\| = 1\},$$

which is a convex and compact subset of $\mathbb{C}$ containing all eigenvalues of A. Further, we make use of the supremum norm

$$\|f\|_{\mathbb{E}} := \sup_{z\in\mathbb{E}} |f(z)|$$

on a subset $\mathbb{E} \subset \mathbb{C}$ for which f is defined. Using these definitions, Crouzeix's theorem states that

$$\|f(A)\| \le C\|f\|_{\mathbb{E}} \tag{4.1}$$

with a constant $C \le 1 + \sqrt{2}$ for any function f which is analytic in a neighborhood of $\mathbb{E} \supseteq W(A)$. Notice that if A is Hermitian then we can compute $\|f(A)\|$ in terms of the eigenvalues of A, and thus obviously (4.1) holds with $C = 1$.

We now use (4.1) together with the polynomial exactness property proven in Theorem 3.2 to obtain bounds on the error at the mth step of our method:

$$E_m(f) := f(A + \boldsymbol{b}\boldsymbol{c}^*) - f(A) - U_m X_m(f) V_m^*. \tag{4.2}$$

THEOREM 4.1. *Let A be Hermitian and let f be defined on a compact set $\mathbb{E}$ containing $W(A) \cup W(A + \boldsymbol{b}\boldsymbol{b}^*)$. Then, the error* (4.2) *with $U_m = V_m$ and $X_m(f)$ from* (2.3) *satisfies*

$$\|E_m(f)\| \le 4 \min_{p\in\Pi_m} \|f - p\|_{\mathbb{E}}.$$

Proof. By the polynomial exactness property of $U_m X_m(f) U_m^*$ stated in Theorem 3.2, we have

$$E_m(f) = E_m(f) - p(A + \boldsymbol{b}\boldsymbol{b}^*) + p(A) + U_m X_m(p) U_m^* = E_m(f-p)$$

for any polynomial $p \in \Pi_m$. For arbitrary $p \in \Pi_m$ we thus have

$$\begin{aligned}\|E_m(f)\| &= \|E_m(f-p)\| \\ &= \|(f-p)(A+\boldsymbol{b}\boldsymbol{b}^*) - (f-p)(A) - U_m X_m(f-p) U_m^*\| \\ &\le \|(f-p)(A+\boldsymbol{b}\boldsymbol{b}^*)\| - \|(f-p)(A)\| - \|U_m X_m(f-p) U_m^*\| \\ &\le 2\|f-p\|_{\mathbb{E}} + \|U_m X_m(f-p) U_m^*\|, \end{aligned} \tag{4.3}$$

where the last inequality follows from (4.1). By (2.2) and (2.3), we have

$$U X_m(f-p) U_m^* = U_m\big((f-p)\big(G_m + \|\boldsymbol{b}\|^2 \boldsymbol{e}_1 \boldsymbol{e}_1^*\big) - (f-p)(G_m)\big) U_m^*.$$

In turn,

$$\|U_m X_m(f-p) U_m^*\| = \|X_m(f-p)\| \le 2\|f-p\|_{\mathbb{E}}, \tag{4.4}$$

where, using $W(G_m) \subseteq W(A)$ and $W(G_m + \|\boldsymbol{b}\| \boldsymbol{e}_1 \boldsymbol{e}_1^*) \subseteq W(A + \boldsymbol{b}\boldsymbol{b}^*)$, we again applied (4.1). Inserting (4.4) into (4.3) and taking the minimum over all $p \in \Pi_m$ completes the proof. ∎

Theorem 4.1 allows to derive convergence bounds from known polynomial approximation results for analytic functions. The obtained convergence rates will typically be exponential for functions which are analytic in a neighborhood of $\mathbb{E}$ and superlinear for entire functions like the exponential.

To extend Theorem 4.1 to the non-Hermitian case, we have to assume that f is analytic on the field of values of the block matrix $\mathcal{A}$ from (2.6).

THEOREM 4.2. *Let* $\mathcal{A} := \begin{bmatrix} A & \boldsymbol{b}\boldsymbol{c}^* \\ 0 & A + \boldsymbol{b}\boldsymbol{c}^* \end{bmatrix}$ *and let* f *be analytic in a neighborhood of a compact set* $\mathbb{E}$ *containing* $W(\mathcal{A})$. *Then, the error* (4.2), *with* U_m, $X_m(f)$, *and* V_m *computed by Algorithm 2, satisfies*

$$\|E_m(f)\| \le 2C \min_{p \in \Pi_m} \|f-p\|_{\mathbb{E}}$$

with a constant $C \le 1 + \sqrt{2}$.

Proof. By Lemma 2.2, the update $f(A + \boldsymbol{b}\boldsymbol{c}^*) - f(A)$ is the (1,2) block of $f(\mathcal{A})$, which we denote by $\big[f(\mathcal{A})\big]_{1,2}$. Letting $W_m = \begin{bmatrix} U_m & 0 \\ 0 & V_m \end{bmatrix}$, we note that the columns of W_m are orthonormal and

$$U_m X_m(f) V_m^* = [W_m\, f(W_m^* \mathcal{A} W_m) W_m^*]_{1,2}$$

holds by the definition of $X_m(f)$.

As in the proof of Theorem 4.1 we use the fact that $E_m(f) = E_m(f-p)$ for any $p \in \Pi_m$. Thus, we obtain for arbitrary $p \in \Pi_m$ that

$$\begin{aligned}\|E_m(f)\| &= \|E_m(f-p)\| \\ &= \|(f-p)(A+\boldsymbol{b}\boldsymbol{c}^*) - (f-p)(A) - U_m X_m(f-p) V_m^*\| \\ &\le \|[(f-p)(\mathcal{A})]_{1,2}\| + \|[W_m\,(f-p)(W_m^* \mathcal{A} W_m) W_m^*]_{1,2}\| \\ &\le \|(f-p)(\mathcal{A})\| + \|(f-p)(W_m^* \mathcal{A} W_m)\| \\ &\le 2C\|f-p\|_{\mathbb{E}} \end{aligned}$$

where the last inequality follows from Crouzeix's theorem, using $W(W_m^* \mathcal{A} W_m) \subseteq W(\mathcal{A}) \subseteq \mathbb{E}$. Taking the minimum over all $p \in \Pi_m$ gives the desired result. ∎

Note that the matrix $\mathcal{A}$ from Theorem 4.2 can be easily block-diagonalized:

$$\mathcal{T}^{-1} \mathcal{A} \mathcal{T} = \begin{bmatrix} A & \\ & A + \boldsymbol{b}\boldsymbol{c}^* \end{bmatrix}, \quad \text{with} \quad \mathcal{T} = \begin{bmatrix} I & I \\ 0 & I \end{bmatrix},$$

with the matrix $\mathcal{T}$ having the very modest 2-norm condition number $\kappa(\mathcal{T}) = \left(\frac{1+\sqrt{5}}{2}\right)^2$. Unfortunately, this does not seem to admit any meaningful conclusion about the numerical range of $\mathcal{A}$. In fact, we are not aware of any tight relationship between $W(\mathcal{A})$ and the numerical ranges of A, $A + \boldsymbol{b}\boldsymbol{c}^*$. Writing

$$\mathcal{A} = \begin{bmatrix} A & 0 \\ 0 & A \end{bmatrix} + \begin{bmatrix} 0 & \boldsymbol{b}\boldsymbol{c}^* \\ 0 & \boldsymbol{b}\boldsymbol{c}^* \end{bmatrix} = \begin{bmatrix} A & 0 \\ 0 & A \end{bmatrix} + \begin{bmatrix} \boldsymbol{b} \\ \boldsymbol{b} \end{bmatrix} [\boldsymbol{0}^*, \boldsymbol{c}^*]$$

we obtain from [29, Section 1.0.1 & Property 1.2.10] the inclusion

$$W(\mathcal{A}) \subseteq W(A) + W(\boldsymbol{u}\boldsymbol{v}^*), \quad \boldsymbol{u} = \begin{bmatrix} \boldsymbol{b} \\ \boldsymbol{b} \end{bmatrix}, \ \boldsymbol{v} = \begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{c} \end{bmatrix}$$

where $+$ refers to the Minkowski sum of sets. In general, the field of values of a rank-one matrix $\boldsymbol{u}\boldsymbol{v}^*$ is an ellipse with focal points 0 and $\boldsymbol{v}^*\boldsymbol{u}$ and minor semi-axis $\frac{1}{2}(\|\boldsymbol{u}\|^2 \|\boldsymbol{v}\|^2 - |\boldsymbol{v}^*\boldsymbol{u}|^2)^{1/2}$, which in our special case amounts to focal points 0 and $\boldsymbol{c}^*\boldsymbol{b}$ and minor semi-axis $\frac{1}{2}(2\|\boldsymbol{b}\|^2 \|\boldsymbol{c}\|^2 - |\boldsymbol{c}^*\boldsymbol{b}|^2)^{1/2}$.

In Section 6.2, we will illustrate for an example of practical relevance that $W(\mathcal{A})$ can have rather undesirable properties, to the extent that Theorem 4.2 is of little help in understanding the convergence of our algorithms. In the next section, we therefore derive convergence bounds by an approach that avoids the dependence on $W(\mathcal{A})$ and only depends on $W(A)$ and $W(A + \boldsymbol{b}\boldsymbol{c}^*)$. While the second set may still be larger than $W(A)$, it is at least always smaller than $W(\mathcal{A})$.

5. Convergence results based on integral representations. We begin by considering results based on the Cauchy integral formula in Section 5.1 and then focus on the special case of Markov functions in Section 5.2.

5.1. Convergence results based on the Cauchy integral formula. Let f be analytic on a domain $\mathbb{E}$ containing the eigenvalues of A and $A+\boldsymbol{b}\boldsymbol{c}^*$. We recall (2.8):

$$f(A + \boldsymbol{b}\boldsymbol{c}^*) - f(A) = -\frac{1}{2\pi i} \int_\Gamma f(z)(zI - A)^{-1} \boldsymbol{b}\boldsymbol{c}^* (zI - A - \boldsymbol{b}\boldsymbol{c}^*)^{-1} \, \mathrm{d}z, \tag{5.1}$$

with $\Gamma = \partial\mathbb{E}$. The integrand in (5.1) involves solutions of shifted linear systems. Letting

$$(zI - A)\boldsymbol{x}(z) = \boldsymbol{b} \text{ and } (zI - A - \boldsymbol{b}\boldsymbol{c}^*)^* \boldsymbol{y}(z) = \boldsymbol{c}, \tag{5.2}$$

we can compactly write the right-hand side of (5.1) as

$$f(A + \boldsymbol{b}\boldsymbol{c}^*) - f(A) = \frac{1}{2\pi i} \int_\Gamma f(z) \boldsymbol{x}(z) \boldsymbol{y}(z)^* \, \mathrm{d}z. \tag{5.3}$$

Now, consider the FOM [38] approximations for (5.2), given by

$$\begin{aligned} \boldsymbol{x}_m(z) &:= \|\boldsymbol{b}\| U_m (zI - G_m)^{-1} \boldsymbol{e}_1, \\ \boldsymbol{y}_m(z) &:= \|\boldsymbol{c}\| V_m (\bar{z}I - H_m - \|\boldsymbol{c}\| \boldsymbol{e}_1 \boldsymbol{b}^* V_m)^{-1} \boldsymbol{e}_1, \end{aligned}$$

where we used the Arnoldi decompositions (2.1) and (2.4). Recalling that $X_m(f)$ is defined as the (1,2) block of

$$f\left(\begin{bmatrix} G_m & \|\boldsymbol{b}\|\|\boldsymbol{c}\|\boldsymbol{e}_1\boldsymbol{e}_1^* \\ 0 & H_m^* + \|\boldsymbol{c}\|V_m^*\boldsymbol{b}\boldsymbol{e}_1^* \end{bmatrix}\right)$$

and using contour integration, as in the proof of Lemma 2.2, we find that

$$U_m X_m(f) V_m^* = -\frac{1}{2\pi i}\int_\Gamma f(z)\boldsymbol{x}_m(z)\boldsymbol{y}_m(z)^*\,\mathrm{d}z \tag{5.4}$$

In other words, the approximation to the update matrix can be interpreted as the integral over outer products of FOM approximations for the shifted linear systems (5.2).

Inserting (5.3) and (5.4) into the definition (4.2) of the error gives

$$\begin{aligned} E_m(f) &= -\frac{1}{2\pi i}\int_\Gamma f(z)\big(\boldsymbol{x}(z)\boldsymbol{y}(z)^* - \boldsymbol{x}_m(z)\boldsymbol{y}_m(z)^*\big)\,\mathrm{d}z \\ &= -\frac{1}{2\pi i}\int_\Gamma f(z)\big(\boldsymbol{x}(z)(\boldsymbol{y}(z)-\boldsymbol{y}_m(z))^* + (\boldsymbol{x}(z)-\boldsymbol{x}_m(z))\boldsymbol{y}_m(z)^*\big)\,\mathrm{d}z. \end{aligned} \tag{5.5}$$

Convergence estimates can now be obtained by taking norms in the contour integral (5.5) and bounding all occurring terms. To do so, we first introduce some notation. In the following, $\mathbb{E}$ is a compact, convex set containing both $W(A)$ and $W(A+\boldsymbol{b}\boldsymbol{c}^*)$. The closed unit disk is denoted by $\mathbb{D}$. Denoting by $\overline{\mathbb{C}} := \mathbb{C}\cup\{\infty\}$ the extended complex plane, let ψ be the conformal mapping from $\overline{\mathbb{C}}\setminus\mathbb{D}$ onto $\overline{\mathbb{C}}\setminus\mathbb{E}$, normalized at infinity such that $\psi(\infty)=\infty$, $\psi'(\infty)>0$. Note that $\psi'(w)$ exists for almost all w on $\partial\mathbb{D}$. We are now in the position to state the following norm bounds on the quantities involved in (5.5).

LEMMA 5.1. *For all $m\geq 2$ and $z=\psi(u)\in\mathbb{C}\setminus\mathbb{E}$ we have*

$$\max\Big\{\frac{\|\boldsymbol{x}(z)\|}{\|\boldsymbol{b}\|},\frac{\|\boldsymbol{x}_m(z)\|}{\|\boldsymbol{b}\|},\frac{\|\boldsymbol{y}(z)\|}{\|\boldsymbol{c}\|},\frac{\|\boldsymbol{y}_m(z)\|}{\|\boldsymbol{c}\|}\Big\} \leq \frac{1}{\operatorname{dist}(z,\mathbb{E})} \leq \frac{|u|/\psi'(\infty)}{(|u|-1)^2}, \tag{5.6}$$

$$\max\Big\{\frac{\|\boldsymbol{x}(z)-\boldsymbol{x}_m(z)\|}{\|\boldsymbol{b}\|},\frac{\|\boldsymbol{y}(z)-\boldsymbol{y}_m(z)\|}{\|\boldsymbol{c}\|}\Big\} \leq \frac{4|u|^{1-m}}{|\psi'(u)|(|u|^2-1)} \leq \frac{4|u|^{-m}}{\operatorname{dist}(z,\mathbb{E})}, \tag{5.7}$$

where $\operatorname{dist}(z,\mathbb{E})$ *denotes the distance of z from $\mathbb{E}$.*

Proof. The first inequality in (5.6) follows for $\boldsymbol{x}(z)$ immediately from the fact that

$$\|\boldsymbol{x}(z)\| = \|(zI-A)^{-1}\boldsymbol{b}\| \leq \|(zI-A)^{-1}\|\,\|\boldsymbol{b}\| \leq \frac{\|\boldsymbol{b}\|}{\operatorname{dist}(z,\mathbb{E})},$$

and analogously for $\boldsymbol{x}_m(z)$, $\boldsymbol{y}(z)$, and $\boldsymbol{y}_m(z)$. For the second inequality in (5.6), we recall a result by Kühnau (see Theorem 3.1 in [44] and its proof):

$$\frac{|u|-1}{1+1/|u|} \leq \frac{\operatorname{dist}(\psi(u),\mathbb{E})}{|\psi'(u)|} \leq \frac{|u|^2-1}{|u|}, \tag{5.8}$$

which, in fact, does not require $\mathbb{E}$ to be convex. By convexity of $\mathbb{E}$, we also have the inequality

$$\Big|\frac{\psi'(u)}{\psi'(\infty)}-1\Big| \leq \frac{1}{|u|^2} \tag{5.9}$$

due to Grötzsch and Golusin; see [30, Section 2]. Combining these two inequalities leads to

$$\frac{1}{\operatorname{dist}(z,\mathbb{E})} \le \frac{1+1/|u|}{|\psi'(u)|(|u|-1)} \le \frac{|u|}{\psi'(\infty)(|u|-1)^2},$$

showing (5.6).

We now turn to the second set (5.7) of inequalities. We will make use of properties of the Faber transform $\mathcal{F}$ of a function analytic in the open unit disk and continuous on the closed unit disk; see [5, 16] for more details. Letting $G(v) = \frac{1}{\psi'(u)(u-v)}$ we obtain

$$\begin{aligned} g(\zeta) := \mathcal{F}(G)(\zeta) &= \frac{1}{2\pi i}\int_{\partial\mathbb{E}} G(\psi^{-1}(\widetilde{\zeta}))\frac{d\widetilde{\zeta}}{\widetilde{\zeta}-\zeta} \\ &= \frac{1}{2\pi i}\int_{|v[=1} G(v)\frac{\psi'(v)\,\mathrm{d}v}{\psi(v)-\zeta} = \frac{1}{\psi(u)-\zeta}, \end{aligned} \tag{5.10}$$

where the last equality follows from the residue theorem; see also [5, Eqns (2.5), (2.7)]. Let P be defined by the formula

$$P(v) = \frac{\overline{u}}{\psi'(u)(|u|^2-1)}\Big(\frac{v}{u}\Big)^{m-1} + \frac{1}{\psi'(u)(v-u)}\Big(\Big(\frac{v}{u}\Big)^{m-1}-1\Big),$$

depending on the parameter u. Then it is straightforward to verify that P is a polynomial of degree at most $m-1$, and that

$$G(v)-P(v) = \frac{1}{\psi'(u)(|u|^2-1)}\frac{v\overline{u}-1}{u-v}\Big(\frac{v}{u}\Big)^{m-1}$$

vanishes at 0 because $m \ge 2$. Thus, with the notation of [5, §2], and $p = \mathcal{F}(P) \in \Pi_{m-1}$ we get that $\mathcal{F}(G-P) = \mathcal{F}_+(G-P) = g-p$, using (5.10). Now Theorem 2.1 in [5], being related to the fundamental work of Crouzeix [12, 13], implies

$$\|(zI-A)^{-1}-p(A)\| = \|(g-p)(A)\| \le 2\max_{|v|=1}|G(v)-P(v)| = \frac{2|u|^{1-m}}{|\psi'(u)|(|u|^2-1)}.$$

Since $W(G_m) \subset W(A)$, the same upper bound is obtained for $\|(zI-G_m)^{-1}-p(G_m)\|$. Because of $p \in \Pi_{m-1}$ and the exactness property (3.4), we have that $p(A)\boldsymbol{b} = \|\boldsymbol{b}\|V_m p(G_m)\boldsymbol{e}_1$ and thus

$$\begin{aligned} \frac{\|\boldsymbol{x}(z)-\boldsymbol{x}_m(z)\|}{\|\boldsymbol{b}\|} &= \frac{\|\boldsymbol{x}(z)-p(A)\boldsymbol{b}-(\boldsymbol{x}_m(z)-\|\boldsymbol{b}\|V_m p(G_m)\boldsymbol{e}_1)\|}{\|\boldsymbol{b}\|} \\ &\le \|(zI-A)^{-1}-p(A)\| + \|(zI-G_m)^{-1}-p(G_m)\| \\ &\le \frac{4|u|^{1-m}}{|\psi'(u)|(|u|^2-1)}, \end{aligned}$$

that is, we have shown the first inequality of (5.7). The second inequality follows from a combination with the second inequality in (5.8). The inequalities in (5.7) involving $\boldsymbol{y}(z)$ instead of $\boldsymbol{x}(z)$ are proven in a completely analogous fashion. □

In order to state our first explicit convergence bound, consider the (compact) level sets $\mathbb{E}_r$, which are defined via their complements as $\mathbb{E}_r = \mathbb{C}\setminus\mathbb{E}_r^c$, where $\mathbb{E}_r^c := \{z \in \mathbb{C}\setminus\mathbb{E} : |\psi^{-1}(z)| > r\}$.

THEOREM 5.2. *Suppose that f is analytic in $\mathbb{E}_R$. Then, for $1 < r < R$*

$$\|E_m(f)\| \le R^{-m-1} \frac{16/\psi'(\infty)}{(1-r/R)(1-1/r)^3} \max_{z\in\Gamma_R} |f(z)|\, \|\boldsymbol{b}\|\, \|\boldsymbol{c}\|,$$

where $\Gamma_R = \partial\mathbb{E}_R$.

Proof. Let $\Gamma = \partial\mathbb{E}_r$. Then we have, by taking norms in (5.5),

$$\|E_m(f)\| \le \frac{1}{2\pi} \int_\Gamma |f(z)| \big(\|\boldsymbol{x}(z)\| \|\boldsymbol{y}(z) - \boldsymbol{y}_m(z)\| + \|\boldsymbol{y}_m(z)\| \|\boldsymbol{x}(z) - \boldsymbol{x}_m(z)\| \big) |\,\mathrm{d}z|. \quad (5.11)$$

Because of the polynomial exactness of the Krylov approximation (2.5) proven in Theorem 3.2, the error of our Krylov approximation is the same for f and $f - p$, for any $p \in \Pi_m$. This allows us to conclude from (5.11) that

$$\|E_m(f)\| \le \frac{\|f-p\|_\Gamma}{2\pi} \int_\Gamma \big(\|\boldsymbol{x}(z)\| \|\boldsymbol{y}(z) - \boldsymbol{y}_m(z)\| + \|\boldsymbol{y}_m(z)\| \|\boldsymbol{x}(z) - \boldsymbol{x}_m(z)\| \big) |\,\mathrm{d}z|. \quad (5.12)$$

Applying the substitution $z = \psi(u)$ with $|u| = r$, we obtain from Lemma 5.1 the upper bound

$$\int_{|u|=r} \big(\|\boldsymbol{x}(\psi(u))\| \|\boldsymbol{y}(\psi(u)) - \boldsymbol{y}_m(\psi(u))\| + \|\boldsymbol{y}_m(\psi(u))\| \|\boldsymbol{x}(\psi(u)) - \boldsymbol{x}_m(\psi(u))\| \big) \frac{|\psi'(u)\,\mathrm{d}u|}{2\pi}$$

$$\le \|\boldsymbol{b}\| \|\boldsymbol{c}\| \int_{|u|=r} \frac{8r^{2-m}}{\psi'(\infty)(r^2-1)(r-1)^2} \frac{|\,\mathrm{d}u|}{2\pi} \le \frac{8r^{-m-1}}{\psi'(\infty)(1-1/r)^3} \|\boldsymbol{b}\| \|\boldsymbol{c}\| \quad (5.13)$$

for the integral in (5.12). Further, we can use a partial Faber sum (cf. [5, Remark 3.3] with $\rho = R/r$) to find the Bernstein-type estimate for the best polynomial approximation on the convex set $\mathbb{E}_r$

$$\min_{p\in\Pi_m} \|f-p\|_\Gamma \le \left(\frac{r}{R}\right)^m \frac{2r}{R-r} \max_{z\in\Gamma_R} |f(z)|. \quad (5.14)$$

Inserting (5.13) and (5.14) into (5.12) then yields the desired result. ∎

Let us illustrate Theorem 5.2 for some particular sets $\mathbb{E}$ and the particular case of the exponential function $f(z) = \exp(z)$. We suppose in the following that $\mathbb{E}$ is convex and symmetric with respect to the real axis. Then $\psi(R) \in \mathbb{R}$ is the element in $\mathbb{E}_R$ with the largest real part, and hence

$$R^{-m-1} \max_{z\in\Gamma_R} |f(z)| = e^{\psi(R)}/R^{m+1}.$$

Our aim will be to choose $R > 1$ to make the above right-hand side small. This task has been (implicitly) accomplished by Hochbruck and Lubich [28, Section 3] for various families of sets like real and purely imaginary intervals, disks, and wedge-shaped sets with a corner at $\psi(1)$, see also the analysis of [5, Section 4] where also general convex domains with corners have been considered.

Our upper bounds will be stated in terms of $\psi(1)$, the element of $\mathbb{E}$ of largest real part, and in terms of $\rho = \psi'(\infty)$, the logarithmic capacity of $\mathbb{E}$ (which is increasing as the set becomes larger). For the four families of sets mentioned above, explicit formulas are known for $\psi(R)$ in terms of $\psi(1), \rho$ and R.

We start with some negative result. By convexity, the function $r \mapsto r\psi'(r)$ is known to be increasing. Provided that $\psi'(1) \ge m+1$, elementary calculus implies

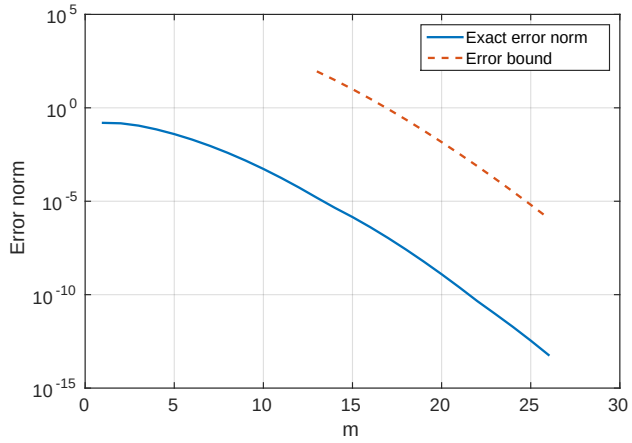

FIG. 5.1. *Exact error norm and error bound* (5.15) *for the update of the matrix exponential described in Example 5.4.*

that $R \mapsto e^{\psi(R)}/R^{m+1}$ is increasing for $R \in [1,\infty)$, and thus Theorem 5.2 is not useful in this case. However, for $m+1 \gg 2\rho \geq \psi'(1)$, we get the following result of superlinear convergence.

COROLLARY 5.3. *Let $\mathbb{E}$ be convex and symmetric with respect to the real axis. Then, under the conditions of Theorem 5.2 and $f(z) = \exp(z)$, we have for $m+1 \geq e\rho$*

$$\|E_m(f)\| \leq \frac{672}{\rho} e^{\psi(1)} \Big(\frac{\rho e}{m+1}\Big)^{m+1} \|\boldsymbol{b}\| \, \|\boldsymbol{c}\|. \tag{5.15}$$

Proof. From the Grötzsch and Golusin inequality (5.9), we find for $1 \leq r \leq R$

$$\psi'(r) - \rho\Big(1 + \frac{1}{r^2}\Big) \leq 0, \text{ where } \rho = \psi'(\infty). \tag{5.16}$$

Integrating (5.16) from 1 to R then yields

$$\psi(R) \leq \psi(1) + \rho\Big(R - \frac{1}{R}\Big).$$

By arguing as in the proof of [28, Theorem 4] we find for $R = \frac{m+1}{\rho}$ that $e^{\psi(R)}/R^{m+1} \leq e^{\psi(1)}(\frac{e\rho}{m+1})^{m+1}$. In addition, since $(1-1/\sqrt{R})^{-4} \leq (1-e^{-1/2})^{-4} \leq 42$, we conclude that

$$\frac{16/\rho}{(1-1/\sqrt{R})^4} \frac{e^{\psi(R)}}{R^{m+1}} \leq \frac{672}{\rho} e^{\psi(1)} \Big(\frac{\rho e}{m+1}\Big)^{m+1},$$

and our claim follows from Theorem 5.2 with $r = \sqrt{R}$. ∎

EXAMPLE 5.4. We illustrate the result of Corollary 5.3 by a simple numerical experiment. We choose $A \in \mathbb{C}^{100\times 100}$ as a diagonal matrix with eigenvalues equidistantly spaced in $[-20, 0]$ and $\boldsymbol{b}$ as a random vector of unit norm, resulting in $\Lambda(A - \boldsymbol{b}\boldsymbol{b}^*) \subseteq [-20.2, 0] =: \mathbb{E}$. We apply Algorithm 1 (modified to account for the update $-\boldsymbol{b}\boldsymbol{b}^*$) for approximating $\exp(A - \boldsymbol{b}\boldsymbol{b}^*) - \exp(A)$ and report the resulting

convergence curve, together with the error bound (5.15) in Figure 5.1. We note that the convergence rate is predicted very accurately, but that the magnitude of the error is severly overestimated due to the large constant in (5.15), which we expect to not be optimal. We thus expect that in general the convergence slope of our bound will be quite accurate, while there is a (large) constant distance between the actual convergence curve and the error bound, something which happens quite frequently for bounds based on the field of values. For practical purposes, we thus suggest to ignore the constant in (5.15).

We remark that by combining the result of Theorem 4.1 with the result of [5, Corollary 4.1] on polynomial approximation of the exponential function, we find a bound that predicts essentially the same convergence rate as the bound (5.15), albeit with a much smaller constant. The real use of the integral representation is thus in the non-Hermitian case, where it allows to circumvent the reliance on $W(\mathcal{A})$. $\diamond$

In view of known results for disks (see [28, Example after Theorem 5]), we do not expect $E_m(f)$ to be small in general for $m+1 < \psi'(1)$. However, in the case of a corner at $\psi(1)$, the right-most element of $\mathbb{E}$, we have a different regime of convergence for $m+1 \ll \rho$. Here we consider only the wedge-like set $\mathbb{E} = \mathbb{E}(\alpha, \rho)$ with

$$\psi(w) = \psi(1) + \rho w\Big(1 - \frac{1}{w}\Big)^{\alpha}, \quad 1 < \alpha \le 2, \tag{5.17}$$

having an outer angle of $\alpha\pi$ at $\psi(1)$, by slightly improving [28, Theorem 6] (though following [5, Corollay 4.2] one could include more general $\mathbb{E}$ having such an angle). Notice that $\mathbb{E}(2,\rho) = [\psi(1) - 4\rho, \psi(1)]$ is a real interval of capacity ρ.

COROLLARY 5.5. *Consider the wedge-like set $\mathbb{E} = \mathbb{E}(\alpha,\rho)$ for $\alpha \in (1,2]$ and $\rho > 0$. Then, under the conditions of Theorem 5.2 and $f(z) = \exp(z)$, we have for $m + 1 - 4/\alpha \in [\alpha\rho^{1/\alpha}, \alpha\rho]$*

$$\|E_m(f)\| \le \frac{(4\rho^{1/\alpha})^4}{\rho} \exp\Big(\psi(1) - (\alpha-1)\Big(\frac{m+1-\frac{4}{\alpha}}{\alpha\rho^{1/\alpha}}\Big)^{\frac{\alpha}{\alpha-1}}\Big)\, \|\boldsymbol{b}\|\, \|\boldsymbol{c}\|. \tag{5.18}$$

Proof. We consider for $R > 1$ the strictly increasing function

$$u = u(R) = \Big(\frac{\psi(R) - \psi(1)}{\rho}\Big)^{1/\alpha} = R^{1/\alpha} - R^{-1+1/\alpha},$$

and notice that $u \le 1$ implies that $R^{1/\alpha} = u + R^{-1+1/\alpha} \le u + 1 \le 2$. Hence, provided that $u \in (0,1)$,

$$\begin{aligned}
\frac{e^{\psi(R)-\psi(1)}}{R^{m+1}(1-1/\sqrt{R})^4} &\le \frac{16 e^{\psi(R)-\psi(1)}}{R^{m+1}(1-1/R)^4} \\
&\le \frac{16}{u^4} \exp\Big(\rho u^{\alpha} + \Big(m+1-\frac{4}{\alpha}\Big)\log\Big(\frac{1}{R}\Big)\Big) \\
&\le \frac{16}{u^4} \exp\Big(\rho u^{\alpha} + \Big(m+1-\frac{4}{\alpha}\Big)\Big(\frac{1}{R}-1\Big)\Big) \\
&= \frac{16}{u^4} \exp\Big(\rho u^{\alpha} - \Big(m+1-\frac{4}{\alpha}\Big)\frac{u}{R^{1/\alpha}}\Big) \\
&\le \frac{16}{u^4} \exp\Big(\rho u^{\alpha} - \Big(m+1-\frac{4}{\alpha}\Big)\frac{u}{2}\Big).
\end{aligned}$$

The argument of the exponential function on the right takes its minimum at

$$u^{\alpha-1} = \frac{m+1-4/\alpha}{2\alpha\rho} \in [\rho^{-\frac{\alpha-1}{\alpha}}, 1]$$

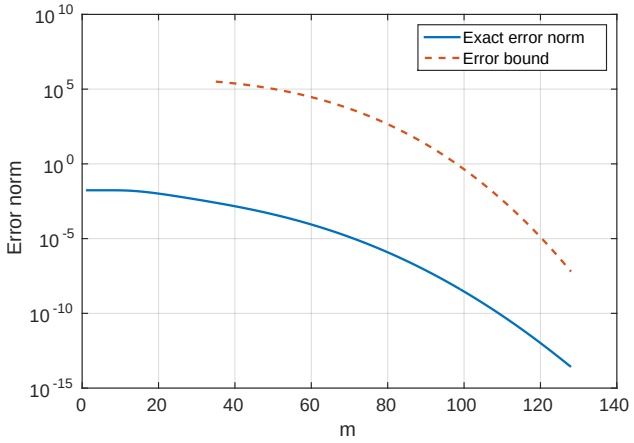

FIG. 5.2. *Exact error norm and error bound* (5.18) *for the update of the matrix exponential described in Example 5.6.*

by our assumption on m. Hence

$$\frac{e^{\psi(R)-\psi(1)}}{R^{m+1}(1-1/\sqrt{R})^4} \le \frac{16}{u^4}\exp\Big(-(\alpha-1)\Big(m+1-\frac{4}{\alpha}\Big)\frac{u}{2\alpha}\Big)$$
$$\le 16\rho^{4/\alpha}\exp\Big(-(\alpha-1)\Big(\frac{m+1-\frac{4}{\alpha}}{2\alpha\rho^{1/\alpha}}\Big)^{\frac{\alpha}{\alpha-1}}\Big),$$

and our claim follows from Theorem 5.2. □

EXAMPLE 5.6. We illustrate the result of Corollary 5.5 for a diagonal matrix $A \in \mathbb{C}^{1000\times 1000}$ with eigenvalues in the wedge-like set (5.17) with $\psi(1) = 0$, $\alpha = 3/2$ and $\rho = 100$. The vector $\boldsymbol{b}$ is again chosen as a random vector of unit norm, which results in $W(A - \boldsymbol{b}\boldsymbol{b}^*) \subseteq \mathbb{E}$ where $\mathbb{E}$ is a wedge-like set corresponding to $\psi(1) = 0$, $\alpha = 1.5$ and $\rho = 101$. We apply Algorithm 2 (modified to account for the update $-\boldsymbol{b}\boldsymbol{b}^*$) for approximating $\exp(A-\boldsymbol{b}\boldsymbol{b}^*)-\exp(A)$ and report the resulting convergence curve together with the error bound (5.18) in Figure 5.2. We again observe that the error norm is overestimated by a few orders of magnitude, while the convergence slope is predicted quite well (although not as accurately as in the Hermitian case in Example 5.4). ◇

In the particular case of symmetric data, $A = A^*$ and $\boldsymbol{c} = \boldsymbol{b}$, the smallest set containing both $W(A)$ and $W(A+\boldsymbol{b}\boldsymbol{b}^*)$ is the interval $\mathbb{E} = [\lambda_{\min}(A), \lambda_{\max}(A+\boldsymbol{b}\boldsymbol{b}^*)]$, where $\lambda_{\min}(\cdot)$ and $\lambda_{\max}(\cdot)$ denote the smallest and largest eigenvalues of a Hermitian matrix. For the exponential function a refined analysis would be possible, using the fact that the coefficients of the Chebyshev series of $f(z) = \exp(z)$ are explicitly known in terms of Bessel functions. We believe, however, that we do not gain much qualitatively compared to the results of the two preceding corollaries with the choices

$$\psi(1) = \lambda_{\max}(A+\boldsymbol{b}\boldsymbol{b}^*), \quad \rho = \frac{\lambda_{\max}(A+\boldsymbol{b}\boldsymbol{b}^*) - \lambda_{\min}(A)}{4}, \quad \alpha = 2.$$

5.2. Convergence results for Markov functions. The approach outlined above extends to other integral representations of f. In particular, a *Markov function*

can be written as

$$f(x) = \int_\alpha^\beta \frac{\mathrm{d}\mu(z)}{x - z}, \tag{5.19}$$

where μ is a positive measure with support in the interval $[\alpha, \beta]$ with $-\infty \leq \alpha < \beta < \infty$. Any such Markov function is analytic in $\overline{\mathbb{C}} \setminus [\alpha, \beta]$. Examples of Markov functions include inverse fractional powers

$$f(z) = z^{-\gamma} = \frac{\sin(\gamma\pi)}{\pi} \int_{-\infty}^0 \frac{(-x)^{-\gamma}\,\mathrm{d}x}{z - x}$$

for $\gamma \in (0, 1)$ or the logarithm

$$f(z) = \frac{1}{z}\log(1 + z) = \int_{-\infty}^{-1} \frac{(-1/x)\,\mathrm{d}x}{z - x},$$

see, e.g., [10, 25] for more details and other examples of Markov functions.

Combining the result of Lemma 5.1 with the integral representation of the error (5.5) now allows to obtain convergence estimates.

THEOREM 5.7. *Let $\mathbb{E}$ be symmetric with respect to the real axis and let $\omega \in \mathbb{E} \cap \mathbb{R}$ denote the element of $\mathbb{E}$ with smallest real part. For a Markov function f with $\alpha < \beta < \omega$ in* (5.19)*, it holds that*

$$\|E_m(f)\| \leq \frac{8\,|f'(\omega)|}{|\phi(\beta)|^m}\,\|\boldsymbol{b}\|\,\|\boldsymbol{c}\|,$$

where ϕ is mapping conformally from $\overline{\mathbb{C}} \setminus \mathbb{E}$ onto $\overline{\mathbb{C}} \setminus \mathbb{D}$.

Proof. Let us first show that we may take $\Gamma = [\alpha, \beta]$ in equation (5.5). Take $r > 1$ such that f is analytic in $\mathbb{E}_r$, that is, $[\alpha, \beta] \cap \mathbb{E}_r = \emptyset$. Since all expressions $\boldsymbol{x}(z), \boldsymbol{y}(z)^*, \boldsymbol{x}_m(z), \boldsymbol{y}_m(z)^*$ are analytic outside $\mathbb{E}$ and decay like $1/z$ at ∞, we get by exchanging integration and using the Cauchy residual theorem

$$\begin{aligned} E_m(f) &= \frac{1}{2\pi i} \int_{\partial\mathbb{E}_r} \int_\alpha^\beta \frac{\mathrm{d}\mu(t)}{t - z} \big(\boldsymbol{x}(z)(\boldsymbol{y}(z) - \boldsymbol{y}_m(z))^* + (\boldsymbol{x}(z) - \boldsymbol{x}_m(z))\boldsymbol{y}_m(z)^*\big)\,\mathrm{d}z \\ &= \int_\alpha^\beta \big(\boldsymbol{x}(t)(\boldsymbol{y}(t) - \boldsymbol{y}_m(t))^* + (\boldsymbol{x}(t) - \boldsymbol{x}_m(t))\boldsymbol{y}_m(t)^*\big)\,\mathrm{d}\mu(t). \end{aligned}$$

By taking norms, we obtain

$$\|E_m(f)\| \leq \int_\alpha^\beta (\|\boldsymbol{x}(t)\|\|\boldsymbol{y}(t) - \boldsymbol{y}_m(t)\| + \|\boldsymbol{y}_m(t)\|\|\boldsymbol{x}(t) - \boldsymbol{x}_m(t))\|)\,\mathrm{d}\mu(t) \tag{5.20}$$

Applying the first inequality in (5.6) and the second inequality in (5.7) to the individual terms in (5.20) then yields

$$\|E_m(f)\| \leq 8\|\boldsymbol{b}\|\|\boldsymbol{c}\| \int_\alpha^\beta \frac{|u|^{-m}\,\mathrm{d}\mu(t)}{\operatorname{dist}(t, \mathbb{E})^2}. \tag{5.21}$$

Using the fact that $t = \psi(u)$, i.e., $u = \phi(t)$ and the fact that $\operatorname{dist}(t, \mathbb{E}) = \omega - t$ for $t \in [\alpha, \beta]$, the inequality (5.21) yields

$$\|E_m(f)\| \leq 8\|\boldsymbol{b}\|\|\boldsymbol{c}\| \int_\alpha^\beta \frac{1}{|\phi(t)|^m} \frac{\mathrm{d}\mu(t)}{(\omega - t)^2}.$$

Since the function $1/|\phi(t)|$ is monotonically increasing on $[\alpha, \beta]$ (see, e.g., the proof of Theorem 6.1 in [5]), we further have

$$\|E_m(f)\| \leq \frac{8\|\boldsymbol{b}\|\|\boldsymbol{c}\|}{|\phi(\beta)|^m} \int_\alpha^\beta \frac{\mathrm{d}\mu(t)}{(\omega - t)^2}. \tag{5.22}$$

The result of the theorem now follows by noting that the integral on the right-hand side of (5.22) is exactly $|f'(\omega)|$, see, e.g., [1]. ∎

We now show how the bound from Theorem 5.7 simplifies when $\mathbb{E}$ is an ellipse on the right of $[\alpha, \beta]$ (or, as a special case, an interval). This gives a more explicit idea of the convergence behavior one can expect.

COROLLARY 5.8. *Let f be a Markov function* (5.19) *with $-\infty \leq \alpha < \beta < \infty$, let $\mathbb{E}$ be an ellipse*

$$\mathbb{E} = \{z \in \mathbb{C} : |z - \sigma + \tau| + |z - \sigma - \tau| \leq \tau(\rho + \rho^{-1})\}, \tag{5.23}$$

where $\sigma \in \mathbb{R}$, $\tau > 0$, $\rho \geq 1$. Let

$$\omega := \sigma - \frac{\tau}{2}(\rho + \rho^{-1}). \tag{5.24}$$

If $\beta < \omega$ and $\mathbb{E}$ is symmetric to the real axis we have

$$\|E_m(f)\| \leq 8\,|f'(\omega)|\,\|\boldsymbol{b}\|\|\boldsymbol{c}\| \left(\rho \cdot \frac{\sqrt{\kappa} - 1}{\sqrt{\kappa} + 1}\right)^m, \quad \kappa = \frac{|\beta - \sigma| + \tau}{|\beta - \sigma| + \tau}. \tag{5.25}$$

Proof. Obviously, ω defined in (5.24) is the element of smallest real part in $\mathbb{E} \cap \mathbb{R}$. The conformal mapping ϕ for $\mathbb{E}$ is given by the inverse Joukowski mapping

$$\phi(z) = \rho^{-1}\left(\zeta + \sqrt{\zeta^2 - 1}\right), \quad \zeta = \frac{z - \sigma}{\tau}.$$

By straight-forward algebraic manipulations, we find that

$$\frac{1}{|\phi(\beta)|} = \rho \frac{\sqrt{\kappa} - 1}{\sqrt{\kappa} + 1}$$

with κ as defined in (5.25). Applying Theorem (5.7) then gives the desired result. ∎

A further simplification is possible in the Hermitian positive definite case, where the ellipse $\mathbb{E}$ is an interval. We assume $\beta \leq 0$ and bound $1/|\phi(\beta)| \leq 1/|\phi(0)|$ in the following result, as this gives a result which closely resembles the classical convergence bound for the conjugate gradient method.

COROLLARY 5.9. *Let f be a Markov function* (5.19) *with $-\infty \leq \alpha < \beta \leq 0$, let A be Hermitian positive definite and let $\boldsymbol{b} = \boldsymbol{c}^*$. Further, let*

$$\kappa_* = \frac{\lambda_{\max}(A + \boldsymbol{b}\boldsymbol{b}^*)}{\lambda_{\min}(A)}$$

. We then have

$$\|E_m(f)\| \leq 8\,|f'(0)|\,\|\boldsymbol{b}\|^2 \left(\frac{\sqrt{\kappa_*} - 1}{\sqrt{\kappa_*} + 1}\right)^m. \tag{5.26}$$

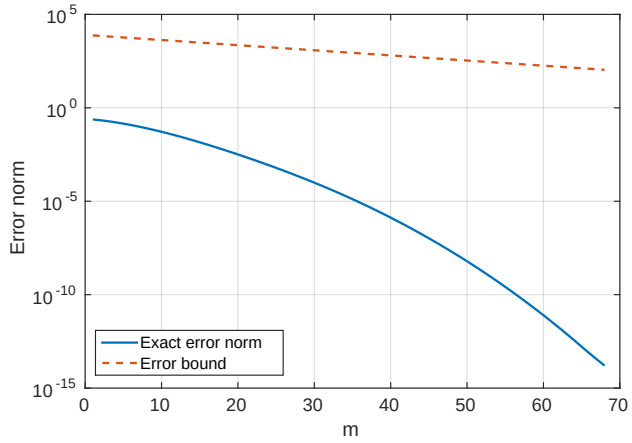

FIG. 5.3. *Exact error norm and error bound* (5.26) *for the update of the matrix inverse square root described in Example 5.10.*

Proof. In the Hermitian positive definite case, $W(A) = [\lambda_{\min}(A), \lambda_{\max}(A)]$ and $W(A + \boldsymbol{b}\boldsymbol{b}^*) = [\lambda_{\min}(A + \boldsymbol{b}\boldsymbol{b}^*), \lambda_{\max}(A + \boldsymbol{b}\boldsymbol{b}^*)]$. As $\lambda_{\min}(A) \leq \lambda_{\min}(A + \boldsymbol{b}\boldsymbol{b}^*)$ and $\lambda_{\max}(A) \leq \lambda_{\max}(A + \boldsymbol{b}\boldsymbol{b}^*)$, we can thus take

$$\mathbb{E} = [\lambda_{\min}(A), \lambda_{\max}(A + \boldsymbol{b}\boldsymbol{b}^*)].$$

This is a special case of an ellipse (5.23) with

$$\rho = 1, \quad \sigma = \tau = \frac{\lambda_{\max}(A + \boldsymbol{b}\boldsymbol{b}^*) - \lambda_{\min}(A)}{2}. \tag{5.27}$$

Noting that $\lambda_{\min}(A)$ is the smallest element in $\mathbb{E} \cap \mathbb{R} = \mathbb{E}$ and inserting (5.27) into (5.25) gives the desired result after some simple calculations. ∎

EXAMPLE 5.10. To illustrate the result of Corollary 5.9, consider a diagonal matrix $A \in \mathbb{R}^{100\times 100}$ with equidistantly spaced eigenvalues in the interval $[0.1, 10]$. As in the previous examples, we choose $\boldsymbol{b}$ as a random vector of unit norm, which results in $\Lambda(A + \boldsymbol{b}\boldsymbol{b}^*) \subseteq [0.1, 10.1] =: \mathbb{E}$. We approximate $(A + \boldsymbol{b}\boldsymbol{b}^*)^{-1/2} - A^{-1/2}$ and report the resulting convergence curve, together with the error bound (5.26) in Figure 5.3. We can observe the typical shortcoming of bounds of the form (5.26) for Markov functions; they only predict linear convergence, while superlinear convergence due to spectral adaption can be observed in this case, see, e.g., [4]. In addition, as we also observed for the exponential function, the large constant in the bound (5.26) leads to a overestimation of the order of magnitude of the error. ◇

REMARK 5.11. We make two remarks concerning Corollary 5.9.

1. Similar to what we pointed out already for the exponential function, if we use known approximation results for Markov functions, like, e.g., [5, Theorem 6.1 & Remark 6.3] together with Theorem 4.1, we obtain the same convergence factor (with a different constant) as in Corollary 5.9 in the Hermitian case.
2. It is possible to obtain a slightly sharper version of (5.26) in which the constant 8 is replaced by 4, and κ_* is replaced by the maximum of the Euclidean norm condition numbers of A and $A + \boldsymbol{b}\boldsymbol{b}^*$ by explicitly exploiting the connection to the conjugate gradient method [26] and integrating over the classical

CG convergence bounds, see, e.g., [20, Lemma 4.1 & Theorem 4.3] for a similar technique. We refrain from giving the details of this, as the improvement over the bound (5.26) is quite marginal in most cases.

6. Applications and numerical experiments. In this section, we present possible applications for the developed methods, with a special focus on the computation of communicability measures in network analysis. All experiments are performed in MATLAB R2016b on a Linux machine with Intel Core i7-6700 CPU and 32 GB main memory.

6.1. Updating network communicability measures. Matrix functions, especially the matrix exponential, play an important role in network analysis, see [18] and the references therein.

Given an undirected graph $G = (V, E)$ with $V = \{1, \dots, n\}$ and $E \subseteq V \times V$, we let $A \in \mathbb{R}^{n \times n}$ be the adjacency matrix of G defined by $a_{ij} = a_{ji} = 1$ if $(i, j) \in E$ and $a_{ij} = 0$ otherwise. Note that A is symmetric. Introduced in [19], the *subgraph centrality* of node i is given by

$$\frac{[\exp(A)]_{ii}}{\operatorname{trace}(\exp(A))}. \tag{6.1}$$

This quantity represents a weighted average of the number of closed walks which connect node i to itself, see, e.g., [18] for details. An interesting question is how the subgraph centralities in a graph change when an edge is added to or removed from the graph; see for example [2], where the effect of adding/removing edges on the so-called *total communicability* is studied. The addition of an edge (i, j), with $i \neq j$, corresponds to the rank-two modification

$$A + BC^*, \quad \text{where} \quad B = [\boldsymbol{e}_i, \boldsymbol{e}_j], \quad C = [\boldsymbol{e}_j, \boldsymbol{e}_i], \tag{6.2}$$

of the adjacency matrix. Analogously, the removal of an edge (i, j) corresponds to a rank-two modification. In turn, these tasks fit perfectly into the framework considered in this paper.

In our experiments:

- Following [7,8], the diagonal entries of $\exp(A)$, needed for evaluating (6.1) for the original matrix A, are estimated by combining the Lanczos method with quadrature [9, 23]. Specifically, we use the implementation of these methods from the `mmq` toolbox [33].
- To handle the rank-two modification (6.2) we perform two consecutive rank-one updates with Algorithm 1, as outlined in Remark 2.3. Note that while $B \neq C$, the resulting matrix BC^* is Hermitian. Thus, we can perform a preprocessing step in order to obtain two Hermitian rank-one updates and then apply Algorithm 1. As said before, it is not necessary to explicitly form the matrix $U_m X_m(\exp) V_m^*$, we only evaluate its diagonal entries.

We now suppose that the diagonal of the matrix exponential has been computed beforehand (in an expensive *offline* calculation). We then add or remove edges from the graph and compare the computation time needed for updating them using Algorithm 1 to the time needed for recomputing them from scratch using the `mmq` toolbox. When using the `mmq` toolbox, we perform five Lanczos iterations per diagonal entry (this value is, e.g., also used in [7]) and can be expected to give a rather rough approximation of the exact value. When using Algorithm 1, we aim for an accuracy of 10^{-6} according to the stopping criterion from Section 2.3, evaluated with $d = 2$. We

TABLE 6.1
Number n of nodes and number k of edges of the networks used in our experiments, together with the computation time for updating the subgraph centrality of all nodes, when modifying 10 edges, using Algorithm 1 vs. recomputing them from scratch using the `mmq` *toolbox.*

Network	n	k	Algorithm 1	`mmq`
Gleich/Minnesota	2,642	6,606	0.15 s	1.80 s
Pajek/Erdos992	6,100	15,030	0.62 s	6.22 s
Pajek/USpowerGrid	4,941	13,188	0.48 s	4.96 s
SNAP/ca-HepTh	9,877	51,971	1.25 s	14.58 s
SNAP/email-Enron	36,692	367,662	1.96 s	147.47 s

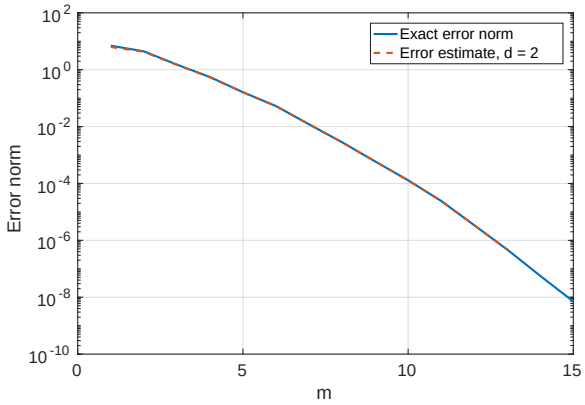

FIG. 6.1. *Convergence curve of Algorithm 2 for computing a rank-one update of the matrix* `Pajek/USpowerGrid` *together with the error estimate* (2.10).

observed that typically between 10 and 30 iterations of the algorithm are necessary to fulfill this criterion. We use five different networks available from the SuiteSparse Matrix Collection [14], and for each network we randomly choose ten edges that are added to or removed from the network (i.e., in order to compute the update, we have to call Algorithm 1 a total of 20 times). Table 6.1 summarizes the obtained results. For all networks under consideration we observe that our algorithm clearly outperforms recomputation.

To also illustrate how the convergence profile of the method looks like and to judge the quality of the error estimate (2.10) in a practical situation, we show the convergence curve of the first update performed for the network `Pajek/USpowerGrid` together with the error estimate in Figure 6.1. We can observe that the method converges very smoothly and that the error estimate is almost identical to the exact error norm for all iterations.

6.2. One-dimensional convection diffusion equation. In the following, we illustrate the convergence of Algorithm 2 for a non-Hermitian example. Consider the one-dimensional convection diffusion equation

$$\begin{aligned} u'' - c \cdot u' &= f \text{ in } [0,1], \\ u(0) = u(1) &= 1, \end{aligned}$$

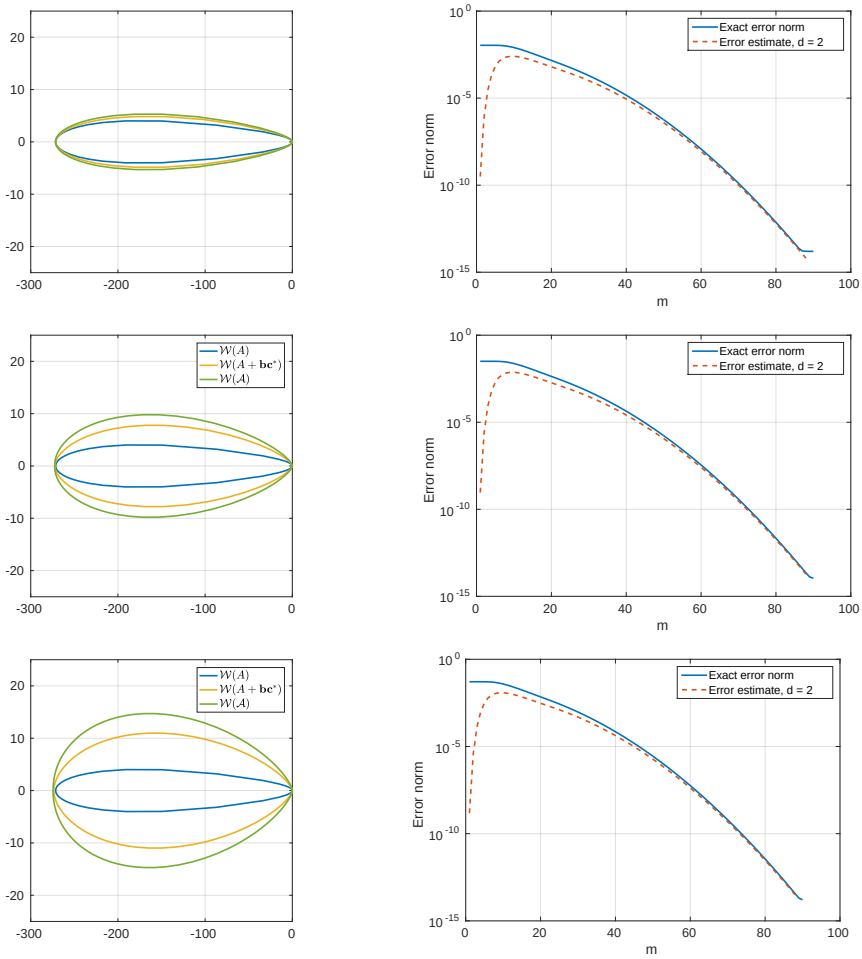

FIG. 6.2. *Field of values (left) for the modifications of the discretized convection diffusion operator corresponding to* $\widetilde{c} = 20, 40, 60$ *(top to bottom) and corresponding convergence curves (right) for approximating* $\exp(A + \boldsymbol{bc}^*) - \exp(A)$.

discretized by centered finite differences combined with a second-order scheme for the convection term. We choose the convection coefficient to be $c = 10$ and discretize the equation on 256 interior grid points. We then consider the rank-one modifications $A + \boldsymbol{bc}^*$ of the discretized operator that change the convection coefficient c in the middle, at the 128th grid point, to (a) $\widetilde{c} = 20$, (b) $\widetilde{c} = 40$, (c) $\widetilde{c} = 60$. On the left-hand side of Figure 6.2, the field of values of A, $A + \boldsymbol{bc}^*$ and the matrix $\mathcal{A}$ from (2.6) is shown for these modifications, and on the right-hand side the convergence history of Algorithm 2 for approximating $\exp(A + \boldsymbol{bc}^*) - \exp(A)$ is given.

As shown in Figure 6.2, the field of values of A lies in a wedge-like set in left half-plane, as considered in Corollary 5.5, with a rather small inner angle at the origin. The angle increases when considering $A + \boldsymbol{bc}^*$ and this increase becomes more pronounced as $\widetilde{c}$ grows. In light of Corollary 5.5, we would thus expect slower convergence for larger values of $\widetilde{c}$. However, as the convergence curves on the right-hand side of Figure 6.2 show, the convergence behavior is virtually identical for all three test cases.

Therefore, our convergence estimates cannot be expected to be sharp in this case. We also report the shape of $W(\mathcal{A})$ in Figure 6.2 in order to illustrate the superiority of the approach from Section 5 over the result of Theorem 4.2, which depends on the field of values of $\mathcal{A}$. Comparing $W(A+\boldsymbol{bc}^*)$ and $W(\mathcal{A})$, we observe that the latter set is substantially larger, especially for large values of $\widetilde{c}$. An especially unfavorable case appears for $\widetilde{c}=60$, where it turns out that $W(\mathcal{A})$ is *not* a subset of the left complex half-plane anymore (and contains 0), unlike $W(A)$ and $W(A+\boldsymbol{bc}^*)$. For an entire function like the exponential function, this only leads to worse convergence bounds. For other functions, like the inverse square root having a singularity at 0, this can be more problematic, to the extent that Theorem 4.2 is not applicable.

Figure 6.2 also shows the difference-based error estimate (2.10) for $d=2$. While this estimate closely follows the exact error curve in later iterations, it severely underestimates the error in the first few iterations. This is due to the fact that the method almost stagnates in these iterations, and is a typical shortcoming of difference-based error estimates.

7. Conclusions. We have proposed Krylov subspace methods for approximating $f(A+D)-f(A)$ for a low-rank matrix D. We proved that the resulting approximation is exact for a polynomial of a certain degree and used this to derive a variety of convergence results, which either link the convergence of our method to polynomial approximation problems or exploit results on the error in the full orthogonalization method in conjunction with an integral representation of f.

In numerical experiments–in particular on applications from network analysis—we have illustrated that our approach can dramatically reduce the cost of computing $f(A+D)$ or portions thereof. We expect that our algorithms will prove useful in other application areas and we will explore this in future work. Another interesting topic for future research is the use of extended and rational Krylov subspaces in the proposed method and an analysis of the resulting convergence behavior.

REFERENCES

[1] H. Alzer and Ch. Berg, *Some classes of completely monotonic functions*, Ann. Acad. Sci. Fenn., Math., 27 (2002), pp. 445–460.

[2] F. Arrigo and M. Benzi, *Updating and downdating techniques for optimizing network communicability*, SIAM J. Sci. Comput., 38 (2016), pp. B25–B49.

[3] A. H. Baker, J. M. Dennis, and E. R. Jessup, *On improving linear solver performance: A block variant of GMRES*, SIAM J. Sci. Comput., 27 (2006), pp. 1608–1626.

[4] B. Beckermann and S. Güttel, *Superlinear convergence of the rational Arnoldi method for the approximation of matrix functions*, Numer. Math., 121 (2012), pp. 205–236.

[5] B. Beckermann and L. Reichel, *Error estimation and evaluation of matrix functions via the Faber transform*, SIAM J. Numer. Anal., 47 (2009), pp. 3849–3883.

[6] P. Benner, P. Ezzatti, D. Kressner, E. S. Quintana-Ortí, and A. Remón, *A mixed-precision algorithm for the solution of Lyapunov equations on hybrid CPU-GPU platforms*, Parallel Comput., 37 (2011), pp. 439–450.

[7] M. Benzi and P. Boito, *Quadrature rule-based bounds for functions of adjacency matrices*, Linear Algebra Appl., 433 (2010), pp. 637–652.

[8] M. Benzi, E. Estrada, and Ch. Klymko, *Ranking hubs and authorities using matrix functions*, Linear Algebra Appl., 438 (2013), pp. 2447–2474.

[9] M. Benzi and G. H. Golub, *Bounds for the entries of matrix functions with applications to preconditioning*, BIT, 39 (1999), pp. 417–438.

[10] Ch. Berg and G. Forst, *Potential Theory on Locally Compact Abelian Groups*, Springer, Berlin Heidelberg, 1975.

[11] D. S. Bernstein and Ch. F. Van Loan, *Rational matrix functions and rank-1 updates*, SIAM J. Matrix Anal. Appl., 22 (2000), pp. 145–154.

[12] M. Crouzeix, *Numerical range and functional calculus in Hilbert space*, J. Funct. Anal., 244 (2007), pp. 668–690.
[13] M. Crouzeix and C. Palencia, *The numerical range as a spectral set*, arxiv:1702.00668, 2017.
[14] T. Davis and Y. Hu, *The SuiteSparse matrix collection.* http://www.cise.ufl.edu/research/sparse/matrices/.
[15] V. Druskin and L. Knizhnerman, *Error bounds in the simple Lanczos procedure for computing functions of symmetric matrices and eigenvalues*, Comput. Math. Math. Phys., 31 (1991), pp. 20–30.
[16] S. W. Ellacott, *On the Faber transform and efficient numerical rational approximation*, SIAM J. Numer. Anal., 20 (1983), pp. 989–1000.
[17] Th. Ericsson, *Computing functions of matrices using Krylov subspace methods*, tech. rep., Department of Computer Science, Chalmers University of Technology and the University of Göteborg, 1990.
[18] E. Estrada and D. J. Higham, *Network properties revealed through matrix functions*, SIAM Rev., 52 (2010), pp. 696–714.
[19] E. Estrada and J. A. Rodríguez-Velázquez, *Subgraph centrality in complex networks*, Phys. Rev. E, 71 (2005), p. 056103.
[20] A. Frommer, S. Güttel, and M. Schweitzer, *Convergence of restarted Krylov subspace methods for Stieltjes functions of matrices*, SIAM J. Matrix Anal. Appl., 35 (2014), pp. 1602–1624.
[21] G. H. Golub and Ch. F. van Loan, *Matrix Computations, 4th edition*, Johns Hopkins University Press, Baltimore, 2013.
[22] G. H. Golub and G. Meurant, *Matrices, moments and quadrature II; How to compute the norm of the error in iterative methods*, BIT, 37 (1997), pp. 687–705.
[23] ———, *Matrices, Moments and Quadrature with Applications*, Princeton University Press, Princeton and Oxford, 2010.
[24] M. H. Gutknecht, *Block Krylov space methods for linear systems with multiple right-hand sides: An introduction*, in Modern Mathematical Models, Methods and Algorithms for Real-World Systems, A. H. Siddiqi, I. S. Duff, and O. Christensen, eds., Anamaya Publishers, New Delhi, 2007.
[25] P. Henrici, *Applied and Computational Complex Analysis, Vol. 2*, John Wiley & Sons, New York, 1977.
[26] M. R. Hestenes and E. Stiefel, *Methods of conjugate gradients for solving linear systems*, J. Res. Natl. Bur. Stand., 49 (1952), pp. 409–436.
[27] N. J. Higham, *Functions of Matrices: Theory and Computation*, SIAM, Philadelphia, 2008.
[28] M. Hochbruck and Ch. Lubich, *On Krylov subspace approximations to the matrix exponential operator*, SIAM J. Numer. Anal., 34 (1997), pp. 1911–1925.
[29] R. A. Horn and Ch. R. Johnson, *Topics in Matrix Analysis*, Cambridge University Press, Cambridge, 1991.
[30] T. Kővari and Ch. Pommerenke, *On Faber polynomials and Faber expansions*, Math. Z., 99 (1967), pp. 193–206.
[31] D. Kressner and C. Tobler, *Krylov subspace methods for linear systems with tensor product structure*, SIAM J. Matrix Anal. Appl., 31 (2010), pp. 1688–1714.
[32] C. Lanczos, *An iteration method for the solution of the eigenvalue problem of linear differential and integral operators*, J. Res. Natl. Stand., 45 (1950), pp. 255–282.
[33] G. Meurant, *MMQ toolbox.* http://pagesperso-orange.fr/gerard.meurant/mmq.zip.
[34] ———, *The computation of bounds for the norm of the error in the conjugate gradient algorithm*, Numerical Algorithms, 16 (1997), pp. 77–87.
[35] G. Meurant and Z. Strakoš, *The Lanczos and conjugate gradient algorithms in finite precision arithmetic*, Acta Numerica, 15 (2006), pp. 471–542.
[36] D. P. O'Leary, *The block conjugate gradient algorithm and related methods*, Linear Algebra Appl., 29 (1980), pp. 293–322.
[37] C. C. Paige, *The computation of eigenvalues and eigenvectors of very large sparse matrices*, PhD thesis, University of London, 1971.
[38] Y. Saad, *Krylov subspace methods for solving large unsymmetric linear systems*, Math. Comput., 37 (1981), pp. 105–126.
[39] Y. Saad, *Analysis of some Krylov subspace approximations to the matrix exponential operator*, SIAM J. Numer. Anal., 29 (1992), pp. 209–228.
[40] H. D. Simon, *Analysis of the symmetric Lanczos algorithm with reorthogonalization methods*, Linear Algebra Appl., 61 (1984), pp. 101–131.
[41] V. Simoncini, *Computational methods for linear matrix equations*, SIAM Rev., 58 (2016), pp. 377–441.

[42] V. Simoncini and E. Gallopoulos, *An iterative method for nonsymmetric systems with multiple right-hand sides*, SIAM J. Sci. Comput., 16 (1995), pp. 917–933.
[43] P. Stange, *Beiträge zu effizienten Algorithmen basierend auf Rang-1 Aufdatierungen*, PhD thesis, Institut Computational Mathematics, TU Braunschweig, 2011.
[44] K.-C. Toh and L. N. Trefethen, *The Kreiss matrix theorem on a general complex domain*, SIAM J. Matrix Anal. Appl., 21 (1999), pp. 145–165.