Subspace extraction for matrix functions ⋆

Marlis Hochbruck [a], Michiel E. Hochstenbach [b,∗]

[a] *Mathematisches Institut, Heinrich-Heine Universität Düsseldorf, Universitätsstraße 1, D-40225 Düsseldorf, Germany*

[b] *Department of Mathematics, Case Western Reserve University, 10900 Euclid Avenue, Cleveland, OH 44106-7058, USA*

Abstract

We investigate existing and novel methods to approximate matrix functions using subspace methods. We analyze a two-sided harmonic Ritz approach and apply this to the extraction from a subspace for matrix functions. We derive all methods in various ways and provide a framework to fit in the techniques.

Key words: Matrix function, matrix exponential, Ritz values, harmonic Ritz values, two-sided Ritz values, two-sided harmonic Ritz values, Petrov values, harmonic Petrov values, interpolation, Dunford–Taylor integral.

1 Introduction: the standard Ritz approach to approximate matrix functions

In the last two decades, there is a vast interest in approximating

$$f(A)\, b, \tag{1.1}$$

where $A \in \mathbb{C}^{n\times n}$ is typically a large sparse matrix and $b \in \mathbb{C}^n$, with Krylov subspace techniques, see, e.g., [1–3, 5, 6, 9, 10, 12, 19, 20]. The best known instance of f, apart from $f(z) = 1/z$ in the linear system context, is the matrix exponential, $f(z) = \exp(z)$, which occurs frequently in the context of the numerical solution of ordinary differential equations, for instance for the use of exponential integrators (see, e.g., [7]). We assume without loss of generality that $\|b\| = 1$, where $\|\cdot\|$ denotes the two-norm. The Krylov relation for $\mathcal{V}_k = \mathcal{K}_k(A, b) = \mathrm{span}\{v_1, Av_1, \ldots, A^{k-1}v_1\}$, where

⋆ This research is supported in part by NSF grant DMS-0405387 and by the Deutsche Forschungsgemeinschaft.

∗ Corresponding author.
Email addresses: `marlis@am.uni-duesseldorf.de` (Marlis Hochbruck), `hochsten@case.edu` (Michiel E. Hochstenbach).

$v_1 = b$ reads

$$AV_k = V_k H_k + \beta_k\, v_{k+1}\, e_k^*, \tag{1.2}$$

where $V_k = [v_1 \ \cdots \ v_k]$ is a basis for $\mathcal{K}_k$, β_k is a short-hand for $h_{k+1,k}$ and e_k is the kth canonical basis vector. Often, but not always, the columns of V_k form an orthonormal basis, which we will assume in this paper unless mentioned otherwise.

By far the most common approximation for x, which we will call the *Ritz approximation*, is (see, e.g., [12])

$$f(A)\, b \approx V_k\, f(H_k)\, e_1. \tag{1.3}$$

We will now review three different derivations of this approximation:

- using a projection onto the search space $\mathcal{V}_k$;
- approximating the shifted linear systems in the Dunford–Taylor integral representation; and
- interpolating the function in certain points.

The first derivation is via the orthogonal projection $V_k V_k^*$ onto the subspace $\mathcal{V}_k$:

$$\begin{aligned} f(A)\, b = f(A) V_k V_k^* b &\approx V_k V_k^* f(A) V_k V_k^* b \\ &\approx V_k f(V_k^* A V_k) V_k^* b = V_k f(H_k)\, e_1. \end{aligned}$$

Secondly, (1.3) can be derived via the Dunford–Taylor integral representation of $f(A)$:

$$f(A)\, b = \frac{1}{2\pi i} \int_\Gamma f(\lambda)(\lambda I - A)^{-1}\, d\lambda\, b, \tag{1.4}$$

where Γ is the boundary curve of a piecewise smooth, bounded region R containing the spectrum of A, assuming that f is analytic in R and continuous on the closure of R; I is the identity matrix. When we solve the shifted linear systems

$$(\lambda I - A)\, x(\lambda) = b, \tag{1.5}$$

using Krylov subspace techniques, then, because of the shift-invariancy $\mathcal{K}_k(\lambda I - A, b) = \mathcal{K}_k(A, b)$ for all A, λ, and b, we can solve all systems simultaneously with the same Krylov space $\mathcal{V}_k = \mathcal{K}_k(A, v_1)$. The Galerkin condition

$$x(\lambda) \approx x_k(\lambda) \in \mathcal{V}_k, \qquad b - (\lambda I - A)\, x_k(\lambda) \perp \mathcal{V}_k,$$

yields

$$x_k(\lambda) = V_k\, (\lambda I - H_k)^{-1}\, e_1.$$

Approximating $x(\lambda) = (\lambda I - A)^{-1} b$ by $x_k(\lambda)$ in (1.4) gives (1.3), if the spectrum of H_k is contained in the region R bounded by Γ, which is for instance the case if R contains the field of values of A.

A third point of view was added by Saad [12]. We recall the following lemma. Let $\mathcal{P}_k$ denote the space of all polynomials of degree $\leq k$. We call (1.2) an Arnoldi-like decomposition if $\dim(V_k) = k$ and H_k is upper Hessenberg; in particular, the columns of V_k do not have to be orthonormal.

Lemma 1 *(see, e.g., [12]) Let A have an Arnoldi-like decomposition $AV_k = V_k H_k + \beta_k v_{k+1} e_k^*$. For all $s \in \mathcal{P}_{k-1}$ we have*

$$s(A)\, b = V_k\, s(H_k)\, e_1.$$

A matrix function $f(A)$ can always be expressed as a matrix polynomial $p_f(A)$, where p_f interpolates f in the eigenvalues (in the Hermite sense). If, in the previous lemma, we choose $s = p_{f,H_k}$ to be the interpolating polynomial of f in the Ritz values of H_k, then we have

$$V_k\, f(H_k)\, e_1 = V_k\, p_{f,H_k}(H_k)\, e_1 = p_{f,H_k}(A)\, b.$$

Therefore, the approximation (1.3) can be seen as an interpolation of f in the Ritz values of A with respect to $\mathcal{V}_k$, that is, the eigenvalues of H_k; this explains the name Ritz approximation. (We note that in the literature, the name *Lanczos* or *Arnoldi approximation* is often used. However, we feel that the name Ritz approximation is more accurate since within the Lanczos or Arnoldi method we can still choose different extraction methods which we will investigate in this paper.)

The rest of this paper is organized as follows. In Section 2 we review and extend the harmonic Ritz extraction for matrix functions. We examine a two-sided harmonic extraction method in Section 3 and use this for an extraction process for matrix functions. In Section 4 we generalize Saad's corrected scheme for all extraction methods and Section 5 provides a unifying framework for all eight derived approximation methods. After some remarks for the Hermitian case in Section 6, we conclude with some numerical illustrations and concluding remarks in Sections 7 and Section 8.

2 Harmonic Ritz approach

An alternative approximation to (1.3) was suggested in Van den Eshof et al. [17,18] for Hermitian matrices in the context of the sign function in a QCD application. This approximation was based on an interpolation in the harmonic Ritz values of A in the target $\tau = 0$. We will now review and generalize this extraction for non-Hermitian A and nonzero target.

The harmonic Ritz extraction was introduced in [11] as an attempt to approximate interior eigenvalues near a target $\tau \in \mathbb{C}$. In practice, a Galerkin extraction often works favorably for exterior eigenvalues but far less favorably for eigenvalues in the interior of the spectrum. Assuming that τ is not an eigenvalue, this suggests to look at the spectrally transformed eigenproblem

$$(A - \tau I)^{-1} x = (\lambda - \tau)^{-1} x. \tag{2.1}$$

Given a search space $\mathcal{V}_k$, we look for approximate eigenpairs (θ, v), $v \in \mathcal{V}_k$, with a (Petrov–)Galerkin condition of the form

$$(A - \tau I)^{-1} v - (\theta - \tau)^{-1} v \perp \mathcal{U}_k.$$

Here, $\mathcal{U}_k$ is a k-dimensional test space which we want to choose such that we avoid working with the inverse of a (presumably large) matrix. This can be done by the test space $\mathcal{U}_k = (A-\tau I)^*(A-\tau I)\,\mathcal{V}_k$. The harmonic Ritz pairs with respect to target τ and search space $\mathcal{V}_k$ are defined as the pairs (θ, v) satisfying the (Petrov–)Galerkin condition

$$(A-\tau I)^{-1}v - (\theta-\tau)^{-1}v \perp (A-\tau I)^*(A-\tau I)\,\mathcal{V}_k$$

or, equivalently,

$$(A-\tau I)\,v - (\theta-\tau)\,v \perp (A-\tau I)\,\mathcal{V}_k.$$

We see that the harmonic Ritz pairs (θ, v) are of the form $\theta = \xi + \tau$, $v = V_k c$, where the (ξ, c) are the eigenpairs of the generalized eigenproblem

$$V_k^*(A-\tau I)^*(A-\tau I)V_k c = \xi\, V_k^*(A-\tau I)^* V_k c. \tag{2.2}$$

Left-multiplying by c^* and using the Cauchy–Schwarz inequality, we get [16]

$$\|(A-\tau I)\,v\| \le |\xi|, \tag{2.3}$$

which indicates that if there is a harmonic Ritz value θ close to τ, the corresponding harmonic Ritz vector has a small residual and hence is an approximate eigenvector of good quality.

Assume that $V_k^*(A-\tau I)^*V_k$ is invertible. Then the harmonic Ritz values are eigenvalues of the matrix

$$\begin{aligned}\widetilde{H}_k &= (V_k^*(A-\tau I)^*V_k)^{-1}\, V_k^*(A-\tau I)^*(A-\tau I)V_k + \tau I\\ &= (V_k^*(A-\tau I)^*V_k)^{-1}\, V_k^*(A-\tau I)^* A V_k.\end{aligned}$$

The harmonic extraction for matrix functions is now defined as the approximation

$$f(A)\,b \approx V_k f(\widetilde{H}_k)\,e_1. \tag{2.4}$$

The idea behind this approximation is that for some functions, a particular target may be important. For instance for the sign function, or any analytic approximation thereof, the target $\tau = 0$ naturally suggests itself. This extraction can, similarly to the Ritz extraction (1.3), be derived in three ways. First, we can use the test space $(A-\tau I)\,\mathcal{V}_k$ to form an oblique projection

$$V_k\,(V_k^*(A-\tau I)^*V_k)^{-1}\, V_k^*(A-\tau I)^*$$

onto $\mathcal{V}_k$ along $((A-\tau I)\,\mathcal{V}_k)^\perp$ so that

$$\begin{aligned}f(A)\,b &\approx V_k\,(V_k^*(A-\tau I)^*V_k)^{-1}\, V_k^*(A-\tau I)^* f(A) V_k\, e_1\\ &\approx V_k\, f((V_k^*(A-\tau I)^*V_k)^{-1}\, V_k^*(A-\tau I)^* A V_k)\, e_1\\ &= V_k f(\widetilde{H}_k)\, e_1.\end{aligned}$$

A second way to derive the harmonic extraction for matrix functions is via the Dunford–Taylor integral (1.4), by approximately solving the associated linear systems (1.5) by the (Petrov–)Galerkin condition

$$x(\lambda) \approx x_k(\lambda) \in \mathcal{V}_k, \qquad b - (\lambda I - A)\, x_k(\lambda) \perp (A - \tau I)\, \mathcal{V}_k.$$

This implies that

$$\begin{aligned} x_k(\lambda) &= [V_k^*(A-\tau I)^*(\lambda I - A)V_k]^{-1}\, V_k^*(A-\tau I)^* V_k e_1 \\ &= [(V_k^*(A-\tau I)^* V_k)^{-1}\, V_k^*(A-\tau I)^*(\lambda I - A)V_k]^{-1}\, e_1 \\ &= [\lambda I - (V_k^*(A-\tau I)^* V_k)^{-1}\, V_k^*(A-\tau I)^* A V_k]^{-1}\, e_1 \\ &= (\lambda I - \widetilde{H}_k)^{-1}\, e_1, \end{aligned}$$

leading to (2.4). Here we have assumed that the region R contains the harmonic Ritz values of A with respect to target τ and search space $\mathcal{V}_k$. (Since the harmonic Ritz values can be infinite, this is not necessarily the case.)

Finally, to interpret (2.4) as an interpolation of f in harmonic Ritz values, we have to use the Krylov relation (1.2):

$$\begin{aligned} \widetilde{H}_k &= (V_k^*(A-\tau I)^* V_k)^{-1}\, V_k^*(A-\tau I)^* A V_k \\ &= (H_k - \tau I)^{-*}[(H_k - \tau I)^* H_k + |\beta_k|^2 e_k e_k^*] \\ &= H_k + |\beta_k|^2 (H_k - \tau I)^{-*} e_k e_k^*. \end{aligned}$$

The harmonic extraction in Krylov context therefore becomes

$$f(A)\, b = V_k\, f(H_k + z_k e_k^*)\, e_1, \qquad z_k = |\beta_k|^2 (H_k - \tau I)^{-*} e_k. \tag{2.5}$$

This approximation is mentioned for Hermitian A (and hence Hermitian tridiagonal H_k) and zero target τ in [18, Equation (21)]. The following lemma is a direct generalization of [18, Lemma 3].

Lemma 2 *For all $s \in \mathcal{P}_{k-1}$ and all $z \in \mathbb{C}^k$ we have*

$$s(A)\, V_k\, e_1 = V_k\, s(H_k + z e_k^*)\, e_1.$$

Proof: The proof by induction on monomials of the form $s(\lambda) = \lambda^j$ uses the fact that $e_k^*(H_k - \tau I)^j e_1 = 0$, for $j = 1, \dots k-1$, for the upper Hessenberg matrix H_k. □

Let $p_{f,\widetilde{H}_k}$ be the Hermite interpolation polynomial of f in the harmonic Ritz values, the eigenvalues of $\widetilde{H}_k$. From Lemma 2, with $z_k = |\beta_k|^2 (H_k - \tau I)^{-*} e_k$, we have

$$V_k\, f(\widetilde{H}_k)\, e_1 = V_k\, p_{f,\widetilde{H}_k}(\widetilde{H}_k)\, e_1 = p_{f,\widetilde{H}_k}(A)\, b.$$

However, we note that we can also reach this interpolation interpretation *directly from Lemma 1*, if we build up an Arnoldi-like relation (1.2) with an $(A - \tau I)$-orthonormal basis V_k, that is,

$$A V_k = V_k \widetilde{H}_k + \widetilde{h}_{k+1,k} v_{k+1} e_k^*, \qquad V_k^*(A - \tau I)^* V_k = I.$$

3 Two-sided harmonic extraction

For non-Hermitian matrices, the main disadvantage of the Arnoldi process is formed by the long recurrences, which means that the amount of work increases per step. Eiermann and Ernst [4] attempt to address this problem by restarting the Arnoldi process for matrix functions.

An attractive alternative to the Arnoldi method is formed by two-sided Lanczos (see, e.g., [21]) because of the short three-term recurrences which read

$$\begin{aligned} Av_k &= \beta_k v_{k+1} + \alpha_k v_k + \gamma_{k-1} v_{k-1}, \\ A^* w_k &= \gamma_k^* w_{k+1} + \alpha_k^* w_k + \beta_{k-1}^* w_{k-1}. \end{aligned}$$

Here we choose $v_1 = b$, and w_1 can be arbitrary such that $w_1^* v_1 = 1$; the α_j, β_j, and γ_j are chosen such that the v_j and w_j are biorthonormal. The corresponding matrix equations are

$$AV_k = V_k T_k + \beta_k v_{k+1} e_k^*, \quad A^* W_k = W_k T_k^* + \gamma_k^* w_{k+1} e_k^*, \quad W_k^* V_k = I_k, \tag{3.1}$$

where the columns of V_k and W_k form biorthonormal basis for $\mathcal{V}_k$ and $\mathcal{W}_k$, respectively. (In fact, V_k and W_k may also be biorthogonal instead of biorthonormal with some appropriate modifications.)

These relations define approximate eigentriples of the form $(\theta, v, w) = (\theta, V_k c, W_k d)$ by the two Galerkin conditions

$$AV_k c - \theta\, V_k c \perp \mathcal{W}_k, \qquad A^* W_k d - \theta\, W_k d \perp \mathcal{V}_k$$

implying that the *two-sided Ritz values* θ and left and right two-sided primitive Ritz vectors c and d are the eigentriples of T_k:

$$W_k^* A V_k c = \theta\, c, \qquad V_k^* A^* W_k d = \theta\, W_k d.$$

(We note that the θs are sometimes also called *Petrov values*.) Although we assume that there is no breakdown, the loss of orthogonality, which typically occurs in finite precision arithmetic, is not something which needs to be feared, see [1] and the numerical experiments in Section 7.

In this section, we first review the two-sided Ritz approximation to matrix functions and give a new property, and then analyze a two-sided harmonic process, which was mentioned in [21], and apply it to the approximation of (1.1).

3.1 Two-sided Ritz

The *two-sided Ritz* approximation

$$f(A)\, b = V_k\, f(T_k)\, e_1, \qquad T_k = W_k^* A V_k \tag{3.2}$$

can be derived in three different ways, similar to Section 1. The first derivation is via the oblique projection $V_kW_k^*$ along the test space $\mathcal{W}_k^\perp$ onto the subspace $\mathcal{V}_k$:

$$\begin{aligned} f(A)\,b = f(A)V_kW_k^*b &\approx V_kW_k^*f(A)V_kW_k^*b \\ &\approx V_kf(W_k^*AV_k)W_k^*b = V_kf(T_k)\,e_1. \end{aligned}$$

The second option is to solve all shifted linear systems (1.5) with a two-sided method using the same Krylov search space $\mathcal{V}_k = \mathcal{K}_k(A, v_1)$ and Krylov test space $\mathcal{W}_k = \mathcal{K}_k(A^*, w_1)$ for all systems ("shifted BiCG"), and impose a (Petrov–)Galerkin condition

$$x(\lambda) \approx x_k(\lambda) \in \mathcal{V}_k, \qquad b - (\lambda I - A)\,x_k(\lambda) \perp \mathcal{W}_k.$$

If the region R contains both the spectra of A and T_k and we insert the resulting $x_k(\lambda) = V_k\,(\lambda I - T_k)^{-1}\,e_1$ for $(\lambda I - A)^{-1}b$ into (1.4), this yields (3.2). A difference with the Ritz approximation is that the spectrum of T_k is not necessarily contained in the field of values of A because of the oblique projection.

Finally, in the same way as Lemma 1, we can show that for all polynomials $s \in \mathcal{P}_{k-1}$

$$s(A)\,b = V_k\,s(T_k)\,e_1.$$

Therefore, the approximation (3.2) can be seen as an interpolation of f in the two-sided Ritz values, the eigenvalues of T_k.

Next, we present a generalization of Saad's theorem [13, Th. 3.6] for two-sided Ritz vectors. Suppose that (θ, v, w) is a two-sided Ritz triple with respect to the search spaces $\mathcal{V}_k$ and $\mathcal{W}_k$, and assume that $w^*v \neq 0$ such that we can scale $w^*v = 1$. Then there exist biorthonormal bases $[v\ V]$ and $[w\ W]$ for $\mathcal{V}_k$ and $\mathcal{W}_k$, which we can expand to biorthonormal bases $[v\ V\ V_\perp]$ and $[w\ W\ W_\perp]$ for $\mathbb{C}^n$. For fixed chosen bases, we have a decomposition

$$x = v(w^*x) + V(W^*x) + V_\perp(W_\perp^*x).$$

If the search and test space are identical ($\mathcal{V} = \mathcal{W}$) and $[v\ V\ V_\perp] = [w\ W\ W_\perp]$ is an orthonormal basis, the quantities

$$c(v, x) := |w^*x|, \qquad s(v, x) := \|[W\ W_\perp]^*x\|, \qquad s(\mathcal{V}, x) := \|W_\perp^*x\|$$

are $\cos(v, x)$, $\sin(v, x)$, and $\sin(\mathcal{V}, x)$, respectively. The projection $[w\ W]^*A[v\ V]$ of A is of the form

$$[w\ W]^*A[v\ V] = \begin{bmatrix} \theta & 0 \\ 0 & G \end{bmatrix} =: \mathrm{diag}(\theta, G), \tag{3.3}$$

where the eigenvalues of G are exactly the two-sided Ritz values with exception of θ. We have the following generalization of Saad's theorem, bounding $s(v, x)$ in terms of $s(\mathcal{V}, x)$.

Theorem 3 *Let (θ, v, w) be a two-sided Ritz triple, $[v\ V\ V_\perp]$ and $[w\ W\ W_\perp]$ be* *biorthonormal bases, and* (λ, x, y) *be an eigentriple. Then*

$$s(v,x) \le s(\mathcal{V},x)\sqrt{1+\frac{\gamma_1^2}{\delta^2}}, \qquad s(w,y) \le s(\mathcal{W},y)\sqrt{1+\frac{\gamma_2^2}{\delta^2}},$$

where

$$\begin{aligned} \gamma_1 &= \|[w\ W]^*\,(A-\lambda I)V_\perp\|, \\ \gamma_2 &= \|W_\perp^*\,(A-\lambda I)[v\ V]\|, \\ \delta &= \mathrm{sep}(\lambda,\, G) := \sigma_{\min}(G-\lambda I) \le \min_{\theta_j \ne \theta} |\theta_j - \lambda|, \end{aligned}$$

and G *is defined as in* (3.3) *and the* θ_j *range over the two-sided Ritz values.*

Proof: Introduce a new variable $z = [z_1^T\ z_2^T\ z_3^T]^T = [w\ W\ W_\perp]^* x$. From $Ax = \lambda x$ we get

$$\begin{bmatrix} w^* \\ W^* \\ W_\perp^* \end{bmatrix} A\,[v\ V\ V_\perp] \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} = \lambda \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix},$$

which we can write as

$$\begin{bmatrix} C & C_1 \\ C_2 & C_3 \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} = \lambda \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix},$$

where $C = \mathrm{diag}(\theta, G)$ and $C_1 = [w\ W]^* A V_\perp$. Therefore we get

$$(C - \lambda I) \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = -C_1 z_3. \tag{3.4}$$

For the right-hand side of (3.4) we have

$$\|C_1 z_3\| \le \|[w\ W]^*(C-\lambda I)V_\perp\|\ \|z_3\|$$

while the left-hand side of (3.4) can be bounded from below by

$$\left\| \begin{bmatrix} \theta-\lambda & 0 \\ 0 & G-\lambda I \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} \right\| \ge \|(G-\lambda I) z_2\| \ge \mathrm{sep}(\lambda, G)\, \|z_2\|.$$

Combining these two bounds, we have

$$\|z_2\| \le \frac{\gamma_1}{\delta}\, \|z_3\|$$

Since $(s(v,x))^2 = \|[W\ W_\perp]^* x\|^2 = \|z_2\|^2 + \|z_3\|^2$ and $\|z_3\| = s(\mathcal{V}, x)$ the result now follows. The "left" version of this theorem is proven similarly. □

The meaning of this theorem is the following. If $\mathcal{V}_k$ "converges towards x" (i.e., $s(\mathcal{V}_k, x)$ is small), and if δ stays away from zero (for this it is sufficient that θ is a simple two-sided Ritz value), then there exists a two-sided Ritz triple (namely, the one with minimal associated $|\theta - \lambda|$) with small $s(v, x)$ (i.e., v converges to x). Two critical remarks: first, the s does not correspond to a sine in the non-Hermitian case; second, the quantities $\gamma_{1,2}$ are not, in contrast to the Hermitian case, bounded by $\|A\|$ because of the oblique projections.

3.2 Two-sided harmonic Ritz

Suppose we are interested in eigenvalues of the eigenvalue problem $Ax = \lambda x$ near a target τ. Again, we consider the transformed eigenproblem (2.1). In view of the two-sided process, we are equally interested in the complex conjugated problem $A^* y = \lambda^* y$, or

$$(A - \tau I)^{-*} y = (\lambda - \tau)^{-*} y.$$

We look for an approximate eigentriple

$$(\theta, v, w) \approx (\lambda, x, y), \qquad v \in \mathcal{V}_k, \quad w \in \mathcal{W}_k,$$

and we would like to use an extraction process that avoids working with the inverse of a matrix and, in addition, extracts approximations to the right and left eigenvector simultaneously.

It turns out that this can be done in the following way. Since $v \in \mathcal{V}_k$ and $w \in \mathcal{W}_k$, we can write $v = V_k c$ and $w = W_k d$. The following four characterizations are equivalent:

$$\begin{array}{llcl}
\text{(i)} & (A - \tau I)^{-1} v - (\theta - \tau)^{-1} v & \perp & ((A - \tau I)^*)^2\, \mathcal{W}_k, \\
& (A - \tau I)^{-*} w - (\theta - \tau)^{-*} w & \perp & (A - \tau I)^2\, \mathcal{V}_k. \\
\text{(ii)} & (A - \tau I)^{-1} \widetilde{v} - (\theta - \tau)^{-1} \widetilde{v} & \perp & \widetilde{\mathcal{W}}_k := (A - \tau I)^*\, \mathcal{W}_k, \quad \widetilde{v} = (A - \tau I)\, v \\
& (A - \tau I)^{-*} \widetilde{w} - (\theta - \tau)^{-*} \widetilde{w} & \perp & \widetilde{\mathcal{V}}_k := (A - \tau I)\, \mathcal{V}_k, \quad \widetilde{w} = (A - \tau I)^* w. \\
\text{(iii)} & (A - \tau I) V_k c - (\theta - \tau) V_k c & \perp & (A - \tau I)^*\, \mathcal{W}_k, \\
& (A - \tau I)^* W_k d - (\theta - \tau)^* W_k d & \perp & (A - \tau I)\, \mathcal{V}_k. \\
\text{(iv)} & W_k^* (A - \tau I)^2 V_k c & = & (\theta - \tau) \quad W_k^* (A - \tau I) V_k c, \\
& d^*\, W_k^* (A - \tau I)^2 V_k & = & (\theta - \tau)\, d^*\, W_k^* (A - \tau I) V_k.
\end{array}$$

Item (iv) indicates a practical procedure: the left and right two-sided harmonic Ritz vectors are left and right eigenvectors of one and the same projected generalized eigenproblem. Assume that $W_k^* (A - \tau I) V_k$ is invertible. We see that the two-sided harmonic Ritz values are the eigenvalues of

$$\begin{aligned}
\widehat{H}_k &:= (W_k^* (A - \tau I) V_k)^{-1} W_k^* (A - \tau I)^2 V_k + \tau I \\
&= (W_k^* (A - \tau I) V_k)^{-1} W_k^* (A - \tau I) A V_k.
\end{aligned} \qquad (3.5)$$

A justification of this approach is the fact that an extract eigentriple (λ, x, y) is also a two-sided harmonic Ritz triple, as may be easily checked.

Proposition 4 *Let* (λ, x, y) *be an eigentriple, then it is also an two-sided harmonic Ritz triple with respect to the subspaces* $\mathcal{V} = \mathrm{span}(x)$ *and* $\mathcal{W} = \mathrm{span}(y)$*.*

The quality of the resulting harmonic Ritz vectors is important and forms the original motivation of introducing harmonic extraction methods. If we select the two-sided harmonic Ritz triple (θ, v, w) of which the value is closest to τ, we may hope that $|\theta - \tau|$ is small. Taking absolute values on both sides of

$$d^* W_k^* (A - \tau I)^2 V_k c = (\theta - \tau)\, d^* W_k^* (A - \tau I) V_k c$$

we get

$$\begin{aligned} \frac{\|(A - \tau I)\, v\|}{\|v\|} &= |\theta - \tau| \left| \frac{\cos(v, (A - \tau I)^* w)}{\cos((A - \tau I)\, v, (A - \tau I)^* w)} \right|, \\ \frac{\|(A - \tau I)^* w\|}{\|w\|} &= |\theta - \tau| \left| \frac{\cos((A - \tau I)\, v, w)}{\cos((A - \tau I)\, v, (A - \tau I)^* w)} \right|. \end{aligned}$$

As in (2.3), we have a relation between the residual norms and the quantity $|\theta - \tau|$. This gives good hope that if there is a two-sided harmonic Ritz value $\theta \approx \tau$, the corresponding two-sided harmonic Ritz vectors have small residual norm and hence are good approximate eigenvectors for an eigenvalue close to τ; unless $(A - \tau I)v$ and $(A - \tau I)^* w$ are almost orthogonal. However, *asymptotically*, that is, when v and w converge to the left and right eigenvector, respectively, we know that both fractions involving the cosines tend to one so this extraction is asymptotically "safe".

If v and w are right and left two-sided harmonic Ritz vectors, and if $w^* v \neq 0$, then we can scale v and w such that $w^* v = 1$. With biorthonormal bases $[v\ V]$ and $[w\ W]$ for $\mathcal{V}_k$ and $\mathcal{W}_k$, we write

$$C := [w\ W]^* (A - \tau I)^2 [v\ V], \qquad B := [w\ W]^* (A - \tau I) [v\ V]. \tag{3.6}$$

From characterization (iv) we have that $C e_1 = (\theta - \tau) B e_1$, so $B^{-1} C$ is of the form

$$B^{-1} C = \begin{bmatrix} \theta - \tau & g^* \\ 0 & \widetilde{G} \end{bmatrix}; \tag{3.7}$$

the eigenvalues of $\widetilde{G}$ are exactly the two-sided harmonic Ritz values except for θ, with τ subtracted. Similarly, we know that $C^* e_1 = (\theta - \tau)^* B^* e_1$, so $B^{-*} C^*$ is of the form

$$B^{-*} C^* = \begin{bmatrix} (\theta - \tau)^* & \widehat{g}^* \\ 0 & \widehat{G}^* \end{bmatrix}; \tag{3.8}$$

the eigenvalues of $\widehat{G}$ are the same as those of $\widetilde{G}$. We have the following counterpart of Theorem 3.

Theorem 5 *Let* (θ, v, w) *be a two-sided* harmonic *Ritz triple,* $[v\ V\ V_\perp]$ *and* $[w\ W\ W_\perp]$ *be biorthonormal bases, and* (λ, x, y) *be an eigentriple. Then*

$$s(v,x) \le s(\mathcal{V},x)\sqrt{1+\frac{\gamma_3^2}{\delta_2^2}\|B^{-1}\|^2}, \qquad s(w,y) \le s(\mathcal{W},y)\sqrt{1+\frac{\gamma_4^2}{\delta_3^2}\|B^{-1}\|^2},$$

where

$$\begin{aligned}
\gamma_3 &= \|[w\ W]^*\,(A-\tau I)(A-\lambda I)V_\perp\|,\\
\gamma_4 &= \|W_\perp^*\,(A-\tau I)(A-\lambda I)[v\ V]\|,\\
\delta_2 &= \operatorname{sep}(\lambda-\tau,\,\widetilde{G}) := \sigma_{\min}(\widetilde{G}-(\lambda-\tau)I) \le \min_{\theta_j\neq\theta}|\theta_j-\lambda|,\\
\delta_3 &= \operatorname{sep}(\lambda-\tau,\,\widehat{G}) := \sigma_{\min}(\widehat{G}-(\lambda-\tau)I) \le \min_{\theta_j\neq\theta}|\theta_j-\lambda|,
\end{aligned}$$

B*,* $\widetilde{G}$*, and* $\widehat{G}$ *are defined as in* (3.6), (3.7), *and* (3.8) *and the* θ_j *range over the two-sided harmonic Ritz values.*

Proof: With the same notation as in the proof of Theorem 3 we have

$$\begin{bmatrix} w^* \\ W^* \\ W_\perp^* \end{bmatrix} (A-\tau I)^2\,[v\ V\ V_\perp] \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} = (\lambda-\tau) \begin{bmatrix} w^* \\ W^* \\ W_\perp^* \end{bmatrix} (A-\tau I)\,[v\ V\ V_\perp] \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix}.$$

which we write in the form

$$\begin{bmatrix} C & C_1 \\ C_2 & C_3 \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} = (\lambda-\tau) \begin{bmatrix} B & B_1 \\ B_2 & B_3 \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix}$$

where $C_1 = [w\ W]^*\,(A-\tau I)^2 V_\perp$ and $B_1 = [w\ W]^*(A-\tau I)V_\perp$. Then we get

$$(B^{-1}C-(\lambda-\tau)I)\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = B^{-1}((\lambda-\tau)B_1 - C_1)z_3. \tag{3.9}$$

The right-hand side of (3.9) is bounded above by

$$\|B^{-1}\|\ \|[w\ W]^*(A-\tau I)(A-\lambda I)V_\perp\|\ \|z_3\|$$

while the left-hand side of (3.9) can be bounded from below using (3.7):

$$\left\|\begin{bmatrix} \theta-\lambda & g^* \\ 0 & \widetilde{G}-(\lambda-\tau)I \end{bmatrix}\begin{bmatrix} z_1 \\ z_2 \end{bmatrix}\right\| \ge \|(\widetilde{G}-(\lambda-\tau)I)z_2\| \ge \operatorname{sep}(\lambda-\tau,\widetilde{G})\,\|z_2\|.$$

Combining these two bounds, we have

$$\|z_2\| \le \frac{\gamma_3}{\delta_2}\, \|B^{-1}\|\, \|z_3\|$$

The rest of the proof is similar to that of Theorem 3. □

3.3 Application to matrix functions

The two-sided harmonic Ritz approximation for matrix function can now be defined as

$$V_k f(\widehat{H}_k)\, e_1, \tag{3.10}$$

with $\widehat{H}_k$ as in (3.5). This approach has the familiar three justifications. The first is seen by (obliquely) projecting

$$\begin{aligned} f(A)\, b &\approx V_k\, (W_k^*(A-\tau I)V_k)^{-1} W_k^*(A-\tau I) f(A) V_k\, e_1 \\ &\approx V_k f((W_k^*(A-\tau I)V_k)^{-1} W_k^*(A-\tau I) A V_k)\, e_1 \\ &= V_k f(\widehat{H}_k)\, e_1. \end{aligned}$$

Secondly, one may check that (3.10) arises if we approximate the shifted linear systems (1.5) by the (Petrov–)Galerkin condition

$$x(\lambda) \approx x_k(\lambda) \in \mathcal{V}_k, \qquad b - (\lambda I - A)\, x_k(\lambda) \perp (A-\tau I)\mathcal{W}_k$$

and the region R contains the eigenvalues and two-sided harmonic Ritz values. For the third derivation, we need the two-sided Krylov relations (3.1). A simple computation learns that

$$\begin{aligned} \widehat{H}_k &= (W_k^*(A-\tau I)V_k)^{-1} W_k^*(A-\tau I) A V_k \\ &= (T_k - \tau I)^{-1} [(T_k - \tau I)\, T_k + \beta_k \gamma_k e_k e_k^*] \\ &= T_k + \beta_k \gamma_k (T_k - \tau I)^{-1} e_k e_k^*; \end{aligned}$$

note in particular that $\widetilde{H}_k$ is a tridiagonal matrix with an extra kth column. We invoke Lemma 2 with $z_k = \beta_k \gamma_k (T_k - \tau I)^{-1} e_k$ to conclude that the approximation (3.10) interpolates f on the two-sided harmonic Ritz values. Alternatively, we use an Arnoldi-like decomposition (1.2) with $(A-\tau I)$-biorthonormal bases:

$$A V_k = V_k \widehat{H}_k + \beta_k v_{k+1} e_k^*, \quad A^* W_k = W_k \widehat{H}_k^* + \gamma_k^* w_{k+1} e_k^*, \quad W_k^*(A-\tau I)V_k = I.$$

and directly use Lemma 1.

4 Corrected schemes

Saad [12] proposed a variant on the Ritz approximation using the "extra" available vector v_{k+1} in the Arnoldi-like decomposition (1.2) for the matrix exponential.

We will formulate this variant for general analytic f, and will show that corrected schemes can also be given for the harmonic and two-sided Ritz approaches. The function

$$\varphi(z) = \frac{f(z) - f(0)}{z}$$

is analytic in the region R and we have

$$f(A)\, b = f(0)\, b + A\, \varphi(A)\, b.$$

Now we can approximate $\varphi(A)\, b$ with any of the four techniques (Ritz, harmonic Ritz, two-sided Ritz, two-sided harmonic Ritz) we have discussed. Let G_k be any of the matrices H_k, $\widetilde{H}_k$, T_k, or $\widehat{H}_k$ for the Ritz, harmonic Ritz, two-sided Ritz, and two-sided harmonic Ritz approximation, respectively. We may then approximate:

$$\begin{aligned} f(A)\, b &\approx f(0)\, b + AV_k \varphi(G_k)\, e_1 \\ &= V_k\, (f(0)\, e_1 + G_k\, \varphi(G_k)\, e_1) + \beta_k\, v_{k+1}\, e_k^* \,\varphi(G_k)\, e_1 \\ &= V_k\, f(G_k)\, e_1 + \beta_k\, v_{k+1}\, e_k^*\, \varphi(G_k)\, e_1, \end{aligned}$$

We get Saad's corrected scheme if we correct the Ritz approach ($G_k = H_k$), cf. [12, Equation (11)], but we can also "correct" the other three methods, i.e., use the extra available Krylov vector.

As a generalization of [12, Prop. 2.1], we can practically compute these corrected approximations as follows. If we define

$$\underline{G}_k = \begin{bmatrix} G_k & 0 \\ \beta_k e_k^* & 0 \end{bmatrix} \tag{4.1}$$

then one may easily check by a Taylor series expansion

$$f(\underline{G}_k) = \begin{bmatrix} f(G_k) & 0 \\ \beta_k e_k^* \varphi(G_k) & f(0) \end{bmatrix},$$

cf. [12, Eq. (13)]. Therefore, we can compute these corrected approximations by

$$V_{k+1} f(\underline{G}_k)\, e_1.$$

We will derive these methods in the three ways mentioned in the introduction. We consider here the corrected Ritz approximation; the other corrected approximation can be derived in a similar way. First, in terms of the orthogonal projector $V_{k+1}V_{k+1}^*$ onto $\mathcal{V}_{k+1}$ we have

$$\begin{aligned} f(A)\, b &\approx V_{k+1}V_{k+1}^* f(A) V_{k+1}\, e_1 \approx V_{k+1} f(V_{k+1}^* A V_{k+1})\, e_1 \\ &\approx V_{k+1} f(V_{k+1}^* A\, [V_k\ \ 0])\, e_1 = V_{k+1} f(\underline{H}_k)\, e_1, \end{aligned}$$

with $\underline{H}_k$ as in (4.1). The inequality on the second line can be seen an extra approximation step.

Alternatively, we get the corrected Ritz approach when we solve the shifted linear systems (1.5) with the Galerkin condition

$$b - (\lambda I - A)\, y_{k+1}(\lambda) \perp \mathcal{V}_{k+1}, \qquad y_{k+1} \in \mathcal{V}_{k+1},$$

perform an extra approximation step

$$\begin{aligned} y_{k+1}(\lambda) &= V_{k+1}\,(\lambda I - V_{k+1}^* A V_{k+1})^{-1}\, e_1 \\ &\approx V_{k+1}\,(\lambda I - V_{k+1}^* A\, [V_k\ 0])^{-1}\, e_1 = V_{k+1}\,(\lambda I - \underline{H}_k)^{-1}\, e_1. \end{aligned}$$

and use this approximation in (1.4), assuming that the region R contains the spectra of both A and $\underline{H}_k$.

Finally, Saad [12] shows that this corrected Ritz approach, which approximates $f(A)$ by $f(0)I + As_{f,H_k}(A)$, where s_{f,H_k} interpolates $\varphi(z)$ on the eigenvalues of H_k, is in fact equivalent with interpolating $f(A)$ in the Ritz values and the additional point zero. This result can be generalized for the other approaches, for instance with $G_k = \widetilde{H}_k$ we interpolate in the harmonic Ritz values plus the point zero.

5 A unifying framework

Summarizing, a unifying framework for the extraction methods treated so far is the following. Let V_k contain a basis for the search space $\mathcal{V}_k$ and let $\mathcal{Y}_k$ be a k-dimensional test space with basis Y_k. If $Y_k^* V_k$ is invertible, then $Z_k^* := (Y_k^* V_k)^{-1} Y_k^*$ satisfies $Z_k^* V_k = I$, while $V_k Z_k^*$ is an oblique projection onto $\mathcal{V}_k$ along $\mathcal{Y}_k^{\perp}$.

The Ritz, harmonic Ritz, two-sided Ritz, and two-sided harmonic Ritz approximations can be derived by

- projecting $f(A)\, b = f(A) V_k Z_k^* b \approx V_k Z_k^* f(A) V_k Z_k^* b \approx V_k f(Z_k^* A V_k) Z^* b$;
- approximating $b - (\lambda I - A)\, x_k(\lambda) \perp \mathcal{Y}_k$; or
- interpolating f on the eigenvalues of $(Y_k^* V_k)^{-1} Y_k^* A V_k$.

Note that only for the interpolation argument we need a Krylov context. Different choices for Y_k lead to different extraction methods, see Table 1.

Table 1
Choice of test space for different extraction methods; the search space is $\mathcal{V}_k$ in all cases.

Method	Test space	Approximation
Ritz	$\mathcal{V}_k$	(1.3)
Harmonic Ritz	$(A - \tau I)\, \mathcal{V}_k$	(2.4)
Two-sided Ritz	$\mathcal{W}_k$	(3.2)
Two-sided harmonic Ritz	$(A - \tau I)^* \mathcal{W}_k$	(3.10)

As we have seen in the previous section, Saad's corrected schemes also fit within this framework, with four small comments:

- the search space is a Krylov space;
- the approximation is sought in $\mathcal{V}_{k+1}$ instead of $\mathcal{V}_k$;
- we employ an extra approximation $V_{k+1}^* A V_{k+1} \approx V_{k+1}^* A\,[V_k\ 0]$;
- the points of interpolation are the (two-sided) (harmonic) Ritz values and the point zero.

6 The Hermitian case

For Hermitian A, with real eigenvalues $\lambda_1 \le \cdots \le \lambda_n$, we can add a different viewpoint as follows, inspired by, for instance, [18]. Let s interpolate f in the nodes $\theta_1, \dots, \theta_k$. For every eigenvalue $\lambda_i \in \mathbb{R}$ we have the expression

$$f(\lambda_i) - s(\lambda_i) = \frac{f^{(k)}(\xi_i)}{k!}\, q(t), \qquad q(t) = \prod_{j=1}^{k} (t - \theta_j),$$

where the $\xi_i \in [\lambda_{\min}, \lambda_{\max}]$. Let G_k be any of the H_k (Ritz), $\widetilde{H}_k$ (harmonic Ritz), T_k (two-sided Ritz), or $\widehat{H}_k$ (two-sided harmonic Ritz) approach. Then for any $s \in \mathcal{P}_{k-1}$ we have

$$f(A)\, b - V_k f(G_k)\, e_1 = (f(A) - s(A))\, b + V_k\, (s(G_k) - f(G_k))\, e_1.$$

Since we can bound [18]

$$\min_{\xi \in [\lambda_1,\, \lambda_n]} \frac{|f^{(k)}(\xi)|}{k!}\, \|q(A)\, b\| \le \|(f(A) - s(A))\, b\| \le \max_{\xi \in [\lambda_1,\, \lambda_n]} \frac{|f^{(k)}(\xi)|}{k!}\, \|q(A)\, b\|,$$

one strategy is to minimize $q(A)\, b$ over all monic polynomials in $\mathcal{P}_{k-1}$. This is done by the so-called *Ritz polynomial* that satisfies

$$q(A)\, b \perp \mathcal{V}_k.$$

Since the Ritz pairs (θ, v) we have

$$(A - \theta I)\, v \perp \mathcal{V}_k,$$

we know that the Ritz values are the zeros of this polynomial.

The harmonic Ritz extraction can be interpreted in a similar way: here we require that

$$q(A)\, b \perp (A - \tau I)\, \mathcal{V}_k.$$

Since for the harmonic Ritz pairs (θ, v) satisfy

$$(A - \theta I)\, v \perp (A - \tau I)\, \mathcal{V}_k,$$

the harmonic Ritz values are the zeros of this *harmonic Ritz polynomial.* If we define the (indefinite) "inner product" $[x,\,y]_{A-\tau I}$ by

$$[x,\,y]_{A-\tau I} := y^*(A-\tau I)x,$$

then the harmonic Ritz approach minimizes the associated "norm" $\|q(A)b\|_{A-\tau I}$. These is a true inner product and norm only if $A-\tau I$ is positive definite, or, alternatively, if $\lambda_1 > \tau$.

7 Numerical examples

We give some typical numerical illustrations. In Figure 1(a) we take a random 1000×1000 upper triangular matrix with $-999,\ldots,0$ as diagonal elements and compute $\exp(A)\,b$, where all elements of b are equal. We test the Ritz (solid), harmonic Ritz ($\tau = 0$, dash), two-sided Ritz (dot) and two-sided harmonic Ritz (dash-dot) methods. The initial left vector w_1 for the two-sided process is random. The Ritz extraction curve is the smoothest, while the other approaches sometimes have very large intermediate peaks because of the oblique projections that cause (two-sided) (harmonic) Ritz values with positive real part. These peaks do not influence the final accuracy of the harmonic approach. We use no reorthogonalization in the two-sided methods; the results with reorthogonalization were similar. The accuracy after 200 steps of the two-sided methods is somewhat less, but if we take the amount of work into account, the two-sided methods, in particular the standard two-sided extraction, are attractive.

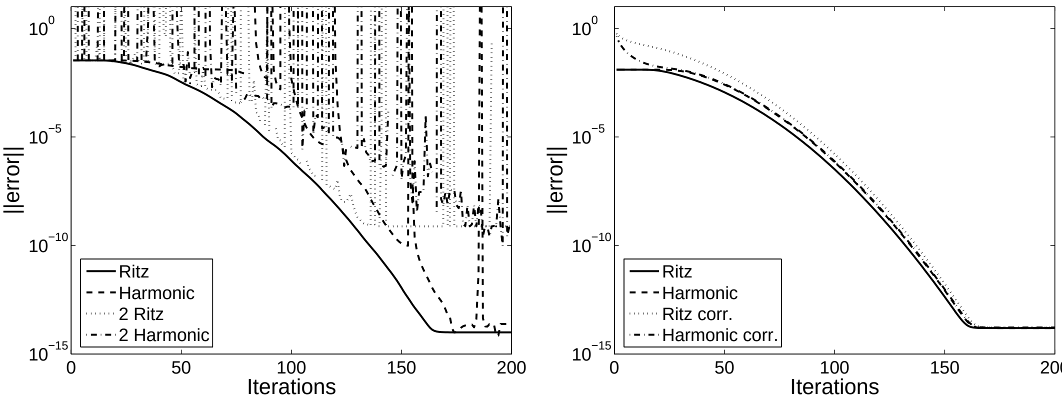

Fig. 1. (a): Converge curves of the Ritz (solid), harmonic Ritz (dashed), two-sided Ritz (dot) and two-sided harmonic Ritz (dash-dot) methods for $f = \exp$ and a 1000×1000 upper triangular matrix with eigenvalues $-999,\ldots,0$. (b): The uncorrected and corrected standard and harmonic Ritz approach for $f = \exp$ and a 1000×1000 diagonal matrix with eigenvalues $-1000,\ldots,-1$.

In Figure 1(b) we again take the matrix exponential of $A = \texttt{diag}(-1000,\ldots,-1)$ and compare the Ritz and harmonic Ritz approach with their "corrected" counterparts. We see that the corrected schemes are similar or slightly *worse* than the uncorrected ones; this corresponds to the fact that the error estimate

$$\beta_k e_k^* \phi(H_k) e_1$$

is often not very accurate, since–in particular in the initial phase of the processes—the Ritz values are not yet accurate approximations to the eigenvalues; cf. also [19].

In Figure 2(a) we take a 1000×1000 diagonal matrix with 500 equidistant eigenvalues in $[-10,-1]$ and 500 in $[1/2, 5]$ and the (non-analytic) function $f = \text{sign}$, which is defined by taking the sign of the eigenvalues. Here, the harmonic extraction with target $\tau = 0$ is more monotonic than the Ritz extraction (see [18] for a heuristic explanation of this phenomenon). Both methods use no reorthogonalization.

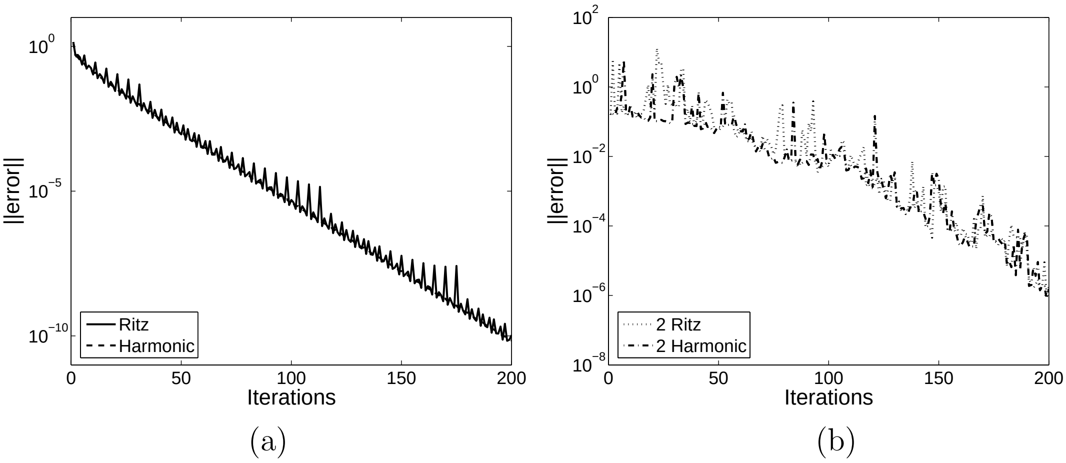

Fig. 2. (a) Convergence curves of the Ritz (solid) and harmonic Ritz (dashed) for $f = \text{sign}$ for a 1000×1000 diagonal matrix with eigenvalues between $[-10,-1]$ and $[1/2, 5]$. (b) Convergence curves of the two-sided Ritz (solid) and to-sided harmonic Ritz (dashed) for $f(z) = 1/z$ for a 1000×1000 upper triangular matrix with eigenvalues between $[-50,-5]$ and $[2, 20]$.

We conclude with the well-known function $f(z) = 1/z$ corresponding to a linear system. We take for A a random upper triangular 1000×1000 matrix with 500 equidistant eigenvalues between $[-50,-5]$ and 500 between $[2, 20]$. We compare the standard two-sided method ("BiCG") and the harmonic two-sided method, for which the approximation $v = Vc$ is determined by the Galerkin condition

$$r = b - AVc \perp A^*\mathcal{W};$$

see Figure 2(b). The two-sided harmonic method has the smallest error in the majority of the iteration steps (about 60%).

8 Discussion and conclusions

In this paper, we investigated existing and novel extraction methods for matrix functions. We analyzed a two-sided harmonic extraction which may also be a useful tool in two-sided methods as two-sided Lanczos (see, e.g., [21]) or two-sided Jacobi–Davidson [8, 15].

We have seen that the extraction from a subspace for matrix functions is relatively well understood in the Hermitian case (see Section 6); for the non-Hermitian case the situation is more subtle. Also in several unreported numerical experiments the convergence of two-sided methods and of one-sided methods was roughly comparable, also without reorthogonalization. This makes the two-sided methods attractive for non-Hermitian matrices. The harmonic approaches may be sensible especially for functions whose interpolation on interior eigenvalues is important. The corrected schemes seem to be less effective in practice.

In the context of eigenvalue problems, it is frequently observed that the harmonic Ritz vectors may be of good quality, but the harmonic Ritz values may be disappointing. Therefore, it has been suggested to take an extra Rayleigh quotient of the harmonic Ritz vectors as new approximate eigenvalues [14]. This is relatively cheap in the sense that in a practical implementation it requires no extra matrix-vector products. This gives the idea of interpolating a matrix function on the Rayleigh quotients of harmonic Ritz vectors, but it is still an open question if this can be done efficiently and effectively.

If the subspace is not a Krylov space we still have the projection and integral interpretation, but we lose the interpolation interpretation, since this required a Krylov relation (1.2). Indeed, from $p(A)v_1 = V_k p(H_k)\, e_1$ for all polynomials with degree $\le k-1$, it follows that for all $j \le k-1$, $A^j b$ must be in $\mathrm{span}(V_k)$, so if the vectors $v_1, \ldots, A^{k-1}v_1$ are independent, then $\mathrm{span}(V_k)$ must be the Krylov space $\mathcal{K}_k(A, b)$.

While this paper dealt with the subspace *extraction*, it is not yet clear how to *expand* a non-Krylov space. We leave this for future work.

References

[1] V. Druskin, A. Greenbaum, and L. Knizhnerman, *Using nonorthogonal Lanczos vectors in the computation of matrix functions*, SIAM J. Sci. Comput., 19 (1998), pp. 38–54.

[2] V. Druskin and L. Knizhnerman, *Krylov subspace approximation of eigenpairs and matrix functions in exact and computer arithmetic*, Numer. Linear Algebra Appl., 2 (1995), pp. 205–217.

[3] ——, *Extended Krylov subspaces: approximation of the matrix square root and related functions*, SIAM J. Matrix Anal. Appl., 19 (1998), pp. 755–771.

[4] E. Eiermann and O. Ernst, *A restarted krylov subspace method for the evaluation of matrix functions*, preprint, Institut für Numerische Mathematik und Optimierung, Technische Universität Bergakademie Freiberg, 2005.

[5] E. Gallopoulos and Y. Saad, *Efficient solution of parabolic equations by Krylov approximation methods*, SIAM J. Sci. Statist. Comput., 13 (1992), pp. 1236–1264.

[6] M. Hochbruck and C. Lubich, *On Krylov subspace approximations to the matrix exponential operator*, SIAM J. Numer. Anal., 34 (1997), pp. 1911–1925.

[7] M. Hochbruck, C. Lubich, and H. Selhofer, *Exponential integrators for large systems of differential equations*, SIAM J. Sci. Comput., 19 (1998), pp. 1552–1574.

[8] M. E. Hochstenbach and G. L. G. Sleijpen, *Two-sided and alternating Jacobi–Davidson*, Linear Algebra Appl., 358(1-3) (2003), pp. 145–172.

[9] C. Moler and C. Van Loan, *Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later*, SIAM Rev., 45 (2003), pp. 3–49.

[10] I. Moret and P. Novati, *Interpolating functions of matrices on zeros of quasi-kernel polynomials*, Numer. Linear Algebra Appl., 12 (2005), pp. 337–353.

[11] C. C. Paige, B. N. Parlett, and H. A. van der Vorst, *Approximate solutions and eigenvalue bounds from Krylov subspaces*, Num. Lin. Alg. Appl., 2 (1995), pp. 115–133.

[12] Y. Saad, *Analysis of some Krylov subspace approximations to the matrix exponential operator*, SIAM J. Numer. Anal., 29 (1992), pp. 209–228.

[13] Y. Saad, *Numerical Methods for Large Eigenvalue Problems*, Manchester University Press, Manchester, UK, 1992.

[14] G. L. G. Sleijpen, H. A. van der Vorst, and E. Meijerink, *Efficient expansion of subspaces in the Jacobi–Davidson method for standard and generalized eigenproblems*, Electron. Trans. Numer. Anal., 7 (1998), pp. 75–89.

[15] A. Stathopoulos, *A case for a biorthogonal Jacobi-Davidson method: restarting and correction equation*, SIAM J. Matrix Anal. Appl., 24 (2002), pp. 238–259.

[16] G. W. Stewart, *Matrix algorithms. Vol. II*, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2001.

[17] J. van den Eshof, *Nested Iteration Methods for Nonlinear Matrix Problems*, PhD thesis, Utrecht University, 2003.

[18] J. van den Eshof, A. Frommer, T. Lippert, K. Schilling, and H. A. van der Vorst, *Numerical methods for the QCD overlap operator: I. Sign-function and error bounds*, Comput. Phys. Comm., 146 (2002), pp. 203–224.

[19] J. van den Eshof and M. Hochbruck, *Preconditioning Lanczos approximations to the matrix exponential*, preprint, Mathematical Institute, Heinrich-Heine Universität Düsseldorf, 2004. To appear in SISC.

[20] H. A. van der Vorst, *An iterative solution method for solving $f(A)x = b$, using Krylov subspace information obtained for the symmetric positive definite matrix A*, J. Comput. Appl. Math., 18 (1987), pp. 249–263.

[21] H. A. van der Vorst, *Iterative Krylov methods for large linear systems*, vol. 13 of Cambridge Monographs on Applied and Computational Mathematics, Cambridge University Press, Cambridge, 2003.