

## SPEEDING UP KRYLOV SUBSPACE METHODS FOR COMPUTING $f(A)b$ VIA RANDOMIZATION\*

ALICE CORTINOVIS<sup>†</sup>, DANIEL KRESSNER<sup>‡</sup>, AND YUJI NAKATSUKASA<sup>§</sup>

**Abstract.** This work is concerned with the computation of the action of a matrix function  $f(A)$ , such as the matrix exponential or the matrix square root, on a vector  $b$ . For a general matrix  $A$ , this can be done by computing the compression of  $A$  onto a suitable Krylov subspace. Such compression is usually computed by forming an orthonormal basis of the Krylov subspace using the Arnoldi method. In this work, we propose to compute (nonorthonormal) bases in a faster way and to use a fast randomized algorithm for least-squares problems to compute the compression of  $A$  onto the Krylov subspace. We present some numerical examples which show that our algorithms can be faster than the standard Arnoldi method while achieving comparable accuracy.

**Key words.** matrix functions, Krylov subspace method, sketching, nonorthonormal basis, randomized algorithms, least-squares problem

**MSC codes.** 65F60, 65F50, 68W20

**DOI.** 10.1137/22M1543458

**1. Introduction.** Given a matrix  $A \in \mathbb{R}^{n \times n}$  and a function  $f$  which is analytic on the closure of a domain  $\mathbb{D} \subset \mathbb{C}$  that contains the eigenvalues of  $A$ , the matrix function  $f(A)$  is defined as

$$f(A) := \frac{1}{2\pi i} \int_{\partial\mathbb{D}} f(z)(zI - A)^{-1} dz,$$

where  $I$  denotes the identity matrix; see the book [17] for equivalent definitions of matrix functions.

Matrix functions arise, for instance, in the solution of partial differential equations [7, 26], in data analysis [5, 12], and in electronic structure calculations [6]. In many of these applications, one is only interested in the product of  $f(A)$  with a vector  $b \in \mathbb{R}^n$ . If  $A$  is large, computing the full  $f(A)$  by standard methods (e.g., the Schur–Parlett algorithm [9]) is infeasible in practice. Moreover, the matrix  $A$  may itself be too large to store explicitly. When  $A$  is only available through matrix-vector multiplications, a Krylov subspace method is often the method of choice to approximate  $f(A)b$  (see, e.g., [17, Chapter 13.2]); we review this approach in section 3.

A key ingredient of Krylov subspace methods is the construction of an orthonormal basis of the Krylov subspace

$$(1.1) \quad \mathcal{K}_m(A, b) = \text{span}\{b, Ab, A^2b, \dots, A^{m-1}b\}$$

\*Received by the editors December 27, 2022; accepted for publication (in revised form) November 14, 2023; published electronically February 9, 2024.

<https://doi.org/10.1137/22M1543458>

**Funding:** The work of the first author has been partially supported by the SNSF research project *Fast Algorithms from Low-Rank Updates*, grant 200020\_178806.

<sup>†</sup>Department of Mathematics, Stanford University, Stanford, CA 94305 USA (alicecor@stanford.edu).

<sup>‡</sup>Institute of Mathematics, EPFL, CH-1015 Lausanne, Switzerland (daniel.kressner@epfl.ch).

<sup>§</sup>Mathematical Institute, University of Oxford, Oxford, OX2 6GG, UK (nakatsukasa@maths.ox.ac.uk).

for some  $m \in \mathbb{N}$  ( $m \ll n$ ), using the Arnoldi algorithm (see, e.g., [23, section 6.2]). However, the cost of orthonormalization grows quadratically with  $m$ , the number of iterations. Many approaches have been proposed to deal with this problem, including explicit polynomial approximation, restarting procedures, and nonsymmetric Lanczos methods; see [15] for a recent survey on this topic.

In this paper, we exploit randomization to speed up Krylov subspace methods for computing  $f(A)b$ . This extends the case of linear systems and eigenvalue problems, which was considered by Balabanov and Grigori in [3] and by Nakatsukasa and Tropp in [21]. The recent work by Güttel and Schweitzer [16] also addresses the computation of  $f(A)b$ ; we will comment in more detail in section 3.3 on its relation to our work.

The key idea of our approach is that there is no need to maintain an orthonormal basis of (1.1) in order to use Krylov subspace methods, as long as one can efficiently compute the compression of  $A$  onto the Krylov subspace. On the one hand, we can exploit faster ways to construct a basis of  $\mathcal{K}_m(A, b)$  (as discussed in [3, 21] and in section 2); on the other hand, we can leverage randomized techniques (such as Blendenpik [2]) in order to compute the compression quickly. We also propose an algorithm that partly bypasses the computation of the compression and is, in turn, even faster—but comes with weaker convergence guarantees.

The rest of this work is organized as follows. In section 2 we review and discuss two algorithms from [3, 21] to construct nonorthonormal bases of  $\mathcal{K}_m(A, b)$ . In section 3 we recall the standard Krylov subspace method and we present our proposed algorithms. In section 4 we present some convergence properties of our methods and discuss their numerical stability. Section 5 contains several numerical examples that illustrate the convergence and timings of the proposed algorithms and a comparison with an algorithm from [16]. Conclusions are summarized in section 6. While our algorithms and results are formulated for real matrices only, most of the discussion carries directly over to complex matrices, under a suitable choice of sketching matrix.

**2. Krylov subspaces and Krylov bases.** In this section, we discuss the computation of an orthonormal basis of  $\mathcal{K}_m(A, b)$  defined in (1.1), the  $m$ th Krylov subspace associated to a matrix  $A \in \mathbb{R}^{n \times n}$  and the vector  $b \in \mathbb{R}^n$ . To simplify the discussion, we will assume that  $\mathcal{K}_m(A, b)$  is an  $m$ -dimensional subspace of  $\mathbb{R}^n$ , which holds whenever  $m < n$  and  $b$  is not contained in a (nontrivial) invariant subspace of  $A$ . An orthonormal basis  $U_m = [u_1 \ \cdots \ u_m]$  of  $\mathcal{K}_m(A, b)$  can be computed using the Arnoldi method (see, e.g., [23, section 6.2]), together with an upper Hessenberg matrix  $K_m \in \mathbb{R}^{m \times m}$  and a unit vector  $u_{m+1}$  such that

$$(2.1) \quad AU_m = U_m K_m + k_{m+1,m} u_{m+1} e_m^T = U_{m+1} \underline{K}_m, \quad \underline{K}_m = \begin{bmatrix} K_m \\ k_{m+1,m} e_m^T \end{bmatrix} \in \mathbb{R}^{(m+1) \times m},$$

where  $e_m$  is the  $m$ th vector of the canonical basis of  $\mathbb{R}^m$ . The relation (2.1) is called the *Arnoldi decomposition*. Note that we have  $U_m^T A U_m = K_m$ , that is,  $K_m$  is the compression of  $A$  onto the  $m$ th Krylov subspace. The Arnoldi method requires  $\mathcal{O}(nm^2)$  operations, in addition to the  $m$  matrix-vector products with  $A$ .

**2.1. Nonorthonormal basis of the Krylov subspace.** As  $m$  increases, the quadratic dependence on  $m$  of the complexity of the Arnoldi method becomes disadvantageous. In the following, we discuss two other ways of constructing bases of  $\mathcal{K}_m(A, b)$ , which are not orthonormal, but can be computed faster. In the works [3, 21] such bases were introduced and used in the context of linear systems and eigenvalue

problems. In our work, we extend this approach to matrix functions; indeed, as we will see in section 3, orthonormality is not necessary for approximating  $f(A)b$ .

The algorithms that we discuss build a basis  $V_m$  of  $\mathcal{K}_m(A, b)$  by rescaling  $b$  and then adding one vector at a time. To enlarge the basis, the last basis vector is premultiplied by  $A$ , and then a suitable linear combination of the previous basis vectors is added to the new vector. A rescaled version of the result becomes the new basis vector. Algorithms that fall into the described framework give rise to an Arnoldi-like decomposition

(2.2)

$$AV_m = V_m H_m + h_{m+1,m} v_{m+1} e_m^T = V_{m+1} \underline{H}_m, \quad \underline{H}_m = \begin{bmatrix} H_m \\ h_{m+1,m} e_m^T \end{bmatrix} \in \mathbb{R}^{(m+1) \times m},$$

where  $V_m$  is the computed basis of  $\mathcal{K}_m(A, b)$  and  $H_m$  is an upper Hessenberg matrix. A key advantage of this more general framework is that it allows for cheaper linear combinations for producing the new basis vector compared to the modified Gram–Schmidt procedure by the Arnoldi method. The two methods from [3, 21] belong to this framework and we summarize them in sections 2.1.1 and 2.1.2, respectively.

**2.1.1. Building the basis by orthogonalizing a sketch.** The idea of the method proposed in [3] is that instead of orthogonalizing each new basis vector against all the previous basis vectors, one maintains a sketched version of the basis and orthogonalizes a sketch of the new vector against the sketch of the previous basis vectors. We now describe more precisely how the basis is constructed, and the procedure is summarized in Algorithm 2.1. First of all we choose a sketch matrix  $\Theta \in \mathbb{R}^{s \times n}$  for some  $m < s \ll n$ . At the  $i$ th step we build a new vector  $w_i$  by multiplying  $A$  with the last vector of the basis. We also keep a sketched version of the basis  $S_{i-1} = \Theta V_{i-1}$ . We compute the sketch  $p_i := \Theta w_i$ , we perform one step of Gram–Schmidt orthogonalization on  $[S_{i-1} \ p_i]$  (which is a small matrix), and we denote by  $S_i$  the resulting  $s \times i$

---

**Algorithm 2.1** Construction of Krylov subspace basis as in [3, Algorithm 2].

---

**Input:** Matrix  $A$ , vector  $b$ , number of iterations  $m$ , parameter  $s \ll n$  (sketch size)

**Output:** Pair  $(V_{m+1}, \underline{H}_m)$  that satisfies Arnoldi-like decomposition (2.2)

- 1: Draw sketch matrix  $\Theta$  of size  $s \times n$
  - 2: Initialize  $R$  as  $(m + 1) \times (m + 1)$  zero matrix
  - 3:  $w_1 \leftarrow b$
  - 4:  $p_1 \leftarrow \Theta w_1$
  - 5:  $r_{11} \leftarrow \|p_1\|_2$
  - 6:  $s_1 \leftarrow p_1/r_{11}$ ,  $S_1 \leftarrow s_1$
  - 7:  $v_1 \leftarrow w_1/r_{11}$ ,  $V_1 \leftarrow v_1$
  - 8: **for**  $i = 2, \dots, m + 1$  **do**
  - 9:    $w_i \leftarrow Av_{i-1}$
  - 10:    $p_i \leftarrow \Theta w_i$
  - 11:    $r_i \leftarrow S_{i-1}^T p_i$ ,  $R(1 : i - 1, i) \leftarrow r_i$
  - 12:    $s_i \leftarrow p_i - S_{i-1} r_i$
  - 13:    $r_{ii} \leftarrow \|s_i\|_2$
  - 14:    $s_i \leftarrow s_i/r_{ii}$ ,  $S_i \leftarrow [S_{i-1} \ s_i]$
  - 15:    $v_i \leftarrow (w_i - V_{i-1} r_i)/r_{ii}$ ,  $V_i \leftarrow [V_{i-1} \ v_i]$
  - 16: **end for**
  - 17:  $\underline{H}_m \leftarrow R(:, 2 : m + 1)$
-

matrix (which has orthonormal columns). We collect the corresponding coefficients  $r_i$  in an upper triangular matrix  $R$ ; see lines 11–14. Then, we use these same coefficients to approximately orthogonalize the vector  $w_i$  with respect to the previous vectors, by computing

$$v_i := (w_i - V_{i-1}r_i)/r_{ii}.$$

The vector  $v_i$  is scaled in such a way that its sketched version has unit length, so that the sketched basis  $S_i$  remains orthonormal. Under suitable assumptions on the sketch matrix  $\Theta$ , the basis  $V_{m+1}$  obtained by Algorithm 2.1 is well-conditioned; see [3, Proposition 4.1].

We choose  $\Theta$  to be a subsampled randomized Hadamard transform (SRHT) [25]:  $\Theta$  consists of selecting (uniformly at random)  $s$  rows from the product of the Walsh–Hadamard transform with a diagonal matrix  $D$  with random independent and identically distributed Rademacher ( $\pm 1$ ) entries on the diagonal. The product of  $\Theta$  with a vector can be computed in  $\mathcal{O}(n \log n)$  time. We remark that other choices are possible, such as the subsampled randomized Fourier transform or sparse sign matrices; see, e.g., [19, sections 8–9].

Let us compare the cost of Algorithm 2.1 with the cost of the standard Arnoldi algorithm. At the  $i$ th iteration, both require a matrix-vector product with  $A$ ; Algorithm 2.1 then requires one multiplication with the current basis  $V_{i-1}$  (in line 15), while the standard Arnoldi algorithm requires two multiplications with  $V_{i-1}$ . All the other computations are comparably inexpensive. Therefore, when, e.g.,  $A$  is very sparse, we can expect Algorithm 2.1 to be up to twice as fast as Arnoldi.

**2.1.2. Building the basis by truncated orthogonalization.** In the following, we describe the “truncated orthogonalization” method proposed in [24, Chapter 6.4.2] and used in [21] in the context of sketching for linear systems and eigenvalue problems. The procedure is summarized in Algorithm 2.2. Each new vector  $w_i = Av_{i-1}$  in line 3 is orthogonalized with respect to the last  $k$  vectors of the basis only, where  $k$  is a fixed small number, e.g.,  $k = 2$ ; see lines 4–10. The matrix  $H_m$  is built iteratively using the coefficients of such a partial orthogonalization. When  $A$  is symmetric, this method coincides for  $k = 2$  with the Lanczos method and gives—in exact arithmetic—an orthonormal basis. When  $A$  is nonsymmetric, the produced basis is not orthogonal in general.

---

**Algorithm 2.2** Construction of basis as in [21] (via truncated orthogonalization).

---

**Input:** Matrix  $A$ , vector  $b$ , number of iterations  $m$ , truncation parameter  $k$

**Output:** Pair  $(V_{m+1}, \underline{H}_m)$  that satisfies Arnoldi-like decomposition (2.2)

```

1:  $v_1 \leftarrow b/\|b\|_2$ 
2: for  $i = 2$  to  $m + 1$  do
3:    $w_i \leftarrow Av_{i-1}$ 
4:   for  $j = \max\{1, i - k\}$  to  $i - 1$  do
5:      $h_{j,i-1} \leftarrow v_j^T w_i$ 
6:      $w_i \leftarrow w_i - h_{j,i} v_j$ 
7:   end for
8:    $h_{i,i-1} \leftarrow \|w_i\|_2$ 
9:    $v_i \leftarrow w_i/\|w_i\|_2$ 
10: end for
11:  $V_{m+1} \leftarrow [v_1, \dots, v_{m+1}]$ 

```

---

The advantage of Algorithm 2.2 is that it avoids the quadratic dependence on  $m$ ; the computational cost is  $\mathcal{O}(nm)$  plus the cost of computing  $m$  matrix-vector multiplications by  $A$ . The main disadvantage, however, is that the basis is prone to become increasingly ill-conditioned and it may actually become numerically rank-deficient already after a few iterations. A potential cure for this behavior is *whitening* (see [21, section 2.2]): One monitors the condition number of the basis (or some related quantity) and once ill-conditioning is observed, an approximate orthogonalization is performed. For this purpose we keep a sketched version of the basis (using a sketch matrix  $\Theta \in \mathbb{R}^{s \times n}$ ), and at the end of each cycle we check whether  $\Theta V_i$  starts to become too ill-conditioned; if this happens, we compute a QR factorization  $\Theta V_i = Q_i R_i$  and consider the basis  $V_i R_i^{-1}$  of  $\mathcal{K}_i(A, b)$  instead, which should now be well-conditioned. Accordingly, the new matrix  $\underline{H}_{i-1}$  in the Arnoldi-like relation (2.2) becomes  $\underline{H}_{i-1} \leftarrow R_i \underline{H}_{i-1} R_i^{-1}$ , where  $R_{i-1}$  is the leading  $(i-1) \times (i-1)$  submatrix of  $R_i$ . Potentially, we could continue the construction of the basis from the pair  $(V_i, \underline{H}_{i-1})$  using Algorithm 2.2 with the updated basis, and perform whitening again whenever it is needed, but we observed that in practice whitening is then needed frequently (every few iterations) and the quality of the basis deteriorates quickly. For this reason, after the first time we whiten the basis, we continue the construction of the basis as in Algorithm 2.1, using the same sketch matrix  $\Theta$ .

*Remark 2.1.* Algorithm 2.1 is mathematically equivalent to Algorithm 2.2 followed by whitening after the last step. However, numerically Algorithm 2.1 is more stable as whitening requires inverting the matrix  $R_{i-1}$  which may be ill-conditioned or even numerically singular.

**3. Computing  $f(A)b$  via Krylov subspace methods.** When  $U_m$  is orthonormal, the classical Krylov subspace projection approach to approximate  $f(A)b$  consists in taking

$$(3.1) \quad f_m := \|b\|_2 U_m f(K_m) e_1, \quad K_m = U_m^T A U_m.$$

The convergence of  $f_m$  to  $f(A)b$  is related to polynomial approximation of  $f$  on a suitable region of the complex plane containing the eigenvalues of  $A$  and will be reviewed at the beginning of section 4.

**3.1. Computing  $f_m$  from Arnoldi-like decompositions.** The approximation (3.1) can be expressed in terms of *any* basis  $V_m$  of the Krylov subspace  $\mathcal{K}_m(A, b)$ . We include the following lemma for completeness.

LEMMA 3.1. *The approximation  $f_m$  defined in (3.1) satisfies*

$$(3.2) \quad f_m = V_m f (V_m^\dagger A V_m) V_m^\dagger b$$

for an arbitrary basis  $V_m$  of  $\mathcal{K}_m(A, b)$ , where  $\dagger$  denotes the pseudoinverse of a matrix.

*Proof.* As  $U_m$  and  $V_m$  span the same subspace, there is an invertible matrix  $R_m \in \mathbb{R}^{m \times m}$  such that  $V_m = U_m R_m$ . Thus,

$$\begin{aligned} V_m f (V_m^\dagger A V_m) V_m^\dagger b &= U_m R_m f (R_m^{-1} U_m^T A U_m R_m) R_m^{-1} U_m^T b \\ &= U_m R_m R_m^{-1} f (U_m^T A U_m) R_m R_m^{-1} U_m^T b \\ &= \|b\|_2 U_m f(K_m) e_1. \end{aligned}$$

□

If an Arnoldi-like decomposition (2.2) is available, like in Algorithms 2.1 and 2.2, the expression (3.2) simplifies. Indeed, we have that

$$(3.3) \quad V_m^\dagger AV_m = H_m + h_{m+1,m} (V_m^\dagger v_{m+1}) e_m^T.$$

Therefore, only the last column of  $V_m^\dagger AV_m$  differs from the matrix  $H_m$  that is already computed. Moreover, we have that  $V_m^\dagger b = \gamma_b e_1$  is a scalar multiple of  $e_1$  with  $\gamma_b = r_{11}$  for Algorithm 2.1 and  $\gamma_b = \|b\|_2$  for Algorithm 2.2.

The vector  $y := V_m^\dagger v_{m+1}$  appearing in (3.3) is the solution of

$$(3.4) \quad y = \arg \min_{z \in \mathbb{R}^m} \|V_m z - v_{m+1}\|_2.$$

When  $n \gg m$ , solving this overdetermined least-squares problem exactly constitutes a significant part of computing the approximation (3.2).

**3.1.1. Solving the least-squares problem.** When the basis  $V_m$  is well-conditioned (e.g., when it is obtained from Algorithm 2.1), a few iterations of LSQR [22] applied to (3.4) are sufficient to get a good approximation of  $V_m^\dagger v_{m+1}$ . When  $V_m$  is not well-conditioned (which is frequently the case when using Algorithm 2.2), LSQR needs to be combined with a preconditioner. For this purpose, we use Blendenpik [2], which computes a random sketch of the basis using, once again, an SRHT, and then uses the  $R$  factor from a QR decomposition of the sketch as a preconditioner. Note that Algorithm 2.1 already orthonormalizes a sketch of the basis (the matrix  $S_m$ ) and therefore the use of Blendenpik, in particular with the same sketch, would be redundant. We remark that it is often not necessary to obtain a very accurate solution of the least-squares problem (3.4); see Lemma 4.2 below. The proposed methods are summarized in Algorithms 3.1 and 3.2.

**3.2. A cheaper approximation.** An alternative to the approximation (3.2) is to evaluate  $f$  on the matrix  $H_m$  instead of  $V_m^\dagger AV_m = H_m + h_{m+1,m} (V_m^\dagger v_{m+1}) e_m^T$ , that is, to approximate

$$(3.5) \quad f(A)b \approx \tilde{f}_m := V_m f(H_m) V_m^\dagger b.$$

This approximation is cheaper to compute because the solution of the least-squares problem is avoided. Note that (3.5) coincides with (3.1) if the basis  $V_m$  is orthonormal,

---

**Algorithm 3.1** Krylov subspace approximation of  $f(A)b$  using Algorithm 2.1.

**Input:** Matrix  $A$ , vector  $b$ , number of iterations  $m$

**Output:** Approximation of  $f(A)b$

- 1: Get Arnoldi-like decomposition  $(V_{m+1}, \underline{H}_m)$  and  $r_{11}$  from Algorithm 2.1
  - 2: Find  $y = \arg \min_{z \in \mathbb{R}^m} \|V_m z - v_{m+1}\|_2$  using LSQR
  - 3: Return  $r_{11} V_m f(H_m + h_{m+1,m} y e_m^T) e_1$
- 

---

**Algorithm 3.2** Krylov subspace approximation of  $f(A)b$  using Algorithm 2.2.

**Input:** Matrix  $A$ , vector  $b$ , number of iterations  $m$

**Output:** Approximation of  $f(A)b$

- 1: Get Arnoldi-like decomposition  $(V_{m+1}, \underline{H}_m)$  from Algorithm 2.2
  - 2: Find  $y = \arg \min_{z \in \mathbb{R}^m} \|V_m z - v_{m+1}\|_2$  using Blendenpik [2]
  - 3: Return  $\|b\|_2 V_m f(H_m + h_{m+1,m} y e_m^T) e_1$
-

---

**Algorithm 3.3**  $f(A)b$  without solving least-squares problems.

---

**Input:** Matrix  $A$ , vector  $b$ , number of iterations  $m$

**Output:** Approximation of  $f(A)b$

- 1: Get Arnoldi-like decomposition  $(V_{m+1}, \underline{H}_m)$  from Algorithm 2.2
  - 2: Return  $\|b\|_2 V_m f(H_m) e_1$
- 

but in general  $H_m \neq V_m^\dagger A V_m$ . We recall that these two Hessenberg matrices only differ in their last column (see (3.3)), and the heuristics behind the approach (3.5) is that this difference may only have a mild influence on the first column of  $f(V_m^\dagger A V_m)$ ; see Theorem 4.3 below for additional theoretical insight. This method is summarized in Algorithm 3.3. Note that for  $f(z) = 1/z$  (that is, when solving a linear system) Algorithm 3.3 becomes equivalent to Saad’s incomplete orthogonalization method described in [24, Algorithm 6.6]. For general  $f$ , an approximation of the form (3.5) for nonorthonormal  $V_m$  is used in the context of restarted methods; in particular Theorem 2.4 in [10] establishes interpolation properties of this approximation.

**3.3. Comparison with related literature:** [16]. While preparing this work, the recent preprint by Güttel and Schweitzer [16] appeared, which proposes— independently from our work—two other randomized methods, called sFOM and sGMRES, for approximating  $f(A)b$ . We focus our comparison on sFOM [16, Algorithm 1], because sGMRES works best for Stieltjes functions and requires numerical quadrature otherwise.

While building a (nonorthonormal) basis  $V_m$  (Algorithm 2.2 is used in the numerical examples of [16]), sFOM also computes  $\Theta V_m$  and  $\Theta A V_m$  for a sketch matrix  $\Theta$ . After computing a QR factorization  $\Theta V_m = S_m R_m$  of the sketched basis, sFOM returns the approximation

$$(3.6) \quad \hat{f}_m := V_m (R_m^{-1} f(S_m^T \Theta A V_m R_m^{-1}) S_m^T \Theta b).$$

If Algorithm 3.3 was combined with the whitening of the basis obtained from Algorithm 2.2 (see section 2.1.2), the returned approximation would be mathematically equivalent to (3.6). To see this, note that the whitened basis of Algorithm 2.2 takes the form  $V_{m+1} R_{m+1}^{-1}$  and therefore the Arnoldi-like relation (2.2) implies

$$(\Theta V_m)^T \Theta A V_m = (\Theta V_m)^T \Theta V_m H_m + h_{m,m+1} (\Theta V_m)^T \Theta v_{m+1} e_{m+1}^T.$$

In turn, as the columns of  $\Theta V_{m+1}$  are orthonormal, this simplifies to  $S_m^T \Theta A V_m = H_m$ . However, the advantage of sFOM over Algorithm 3.3 combined with whitening is that in (3.6) it is not needed to explicitly compute  $V_m R_m^{-1}$ . For both Algorithm 3.3 and sFOM, it is difficult to derive meaningful a priori convergence statements; see Theorem 4.3 and Corollary 2.1 in [16], respectively. In practice, both algorithms can be quite fast in some cases but in other cases their observed convergence behavior is erratic; see Figures 3, 4, and 7 below.

*Remark 3.2.* Another alternative view of sFOM is that it is mathematically equivalent to use Algorithm 2.2 for generating the basis, followed by obtaining the last column of  $V_m^\dagger A V_m$  through solving the least-squares problem  $\arg \min_{z \in \mathbb{R}^m} \|V_m z - v_{m+1}\|_2$  approximately with a sketch-and-solve approach, that is,

$$\arg \min_{z \in \mathbb{R}^m} \|\Theta(V_m z - v_{m+1})\|_2.$$

TABLE 1

*Cost comparison. Arithmetic cost of approximately computing  $f(A)b$  using an  $m$ -dimensional Krylov space, using Algorithms 3.1, 3.2, 3.3, Arnoldi, or sFOM. We assume that the truncation parameter  $k$  in Algorithm 2.2 satisfies  $k \ll m \ll n$  and we denote by  $T_{\text{matvec}}$  the cost of one matrix-vector multiplication with  $A$ .*

	Matrix access	Form $V_m$	Sketch	$f(H_m)$	Form approx.
Arnoldi	$mT_{\text{matvec}}$	$nm^2$	—	$m^3$	$nm$
Algorithm 3.1	$mT_{\text{matvec}}$	$nm^2$	—	$m^3$	$nm$
Algorithm 3.2	$mT_{\text{matvec}}$	$nmk$	$nm \log m$	$m^3$	$nm$
Algorithm 3.3	$mT_{\text{matvec}}$	$nmk$	—	$m^3$	$nm$
sFOM	$mT_{\text{matvec}}$	$nmk$	$nm \log m$	$m^3$	$nm$

Table 1 compares the computational complexity of all algorithms discussed in this section. This reiterates the point that the  $\mathcal{O}(nm^2)$  complexity can be avoided when using Algorithm 2.2.

**4. Convergence analysis.** In this section, we analyze the convergence of Algorithms 3.1, 3.2, and 3.3. For this purpose, let  $\Pi_m$  denote the vector space of polynomials of degree at most  $m$  and let  $W(A) = \{z^*Az \mid z \in \mathbb{C}^n, \|z\|_2 = 1\}$  denote the numerical range of  $A$ . Given a set  $\mathbb{D} \subseteq \mathbb{C}$ , we define  $\|f\|_{\mathbb{D}} := \sup\{|f(z)| : z \in \mathbb{D}\}$ .

**4.1. Analysis of Algorithms 3.1 and 3.2.** We have the following well-known result that links the quality of the approximation (3.2) with polynomial approximation of  $f$  on the numerical range of  $A$ .

**THEOREM 4.1.** *Let  $A \in \mathbb{R}^{n \times n}$ ,  $b \in \mathbb{R}^n$ , and a function  $f : \mathbb{D} \rightarrow \mathbb{C}$  be holomorphic on a domain  $\mathbb{D} \subseteq \mathbb{C}$  containing  $W(A)$ . Then the approximation  $f_m$  from (3.2) satisfies*

$$\|f(A)b - f_m\|_2 \leq 2(1 + \sqrt{2}) \min_{p \in \Pi_{m-1}} \|f - p\|_{W(A)}.$$

*Proof.* This is a classical result for the approximation given by the Arnoldi decomposition; see, e.g., [14, Corollary 3.4]. Thanks to Lemma 3.1, the result extends to any approximation of the form (3.2).  $\square$

Theorem 4.1 assumes that the least-squares problem (3.4) is solved exactly. In line 2 of Algorithms 3.1 and 3.2, only an approximation by an iterative solver is used. The following lemma quantifies the impact of this approximation.

**LEMMA 4.2.** *Let  $\tilde{y}$  be an approximation to  $y = \arg \min_{z \in \mathbb{R}^m} \|V_m z - v_{m+1}\|_2$ . Assume that  $f$  is holomorphic on a convex open set  $\mathbb{D}$  containing  $W(V_m^\dagger AV_m) \cup W(H_m + h_{m+1,m} \tilde{y} e_m^T)$ . Then*

$$\|f(V_m^\dagger AV_m) - f(H_m + h_{m+1,m} \tilde{y} e_m^T)\|_F \leq (1 + \sqrt{2})^2 \|f'\|_{\mathbb{D}} \|h_{m+1,m}(y - \tilde{y})\|_2,$$

where  $\|\cdot\|_F$  denotes the Frobenius norm.

*Proof.* Corollary 4.2 in [4] states that for any square matrices  $C$  and  $D$  of the same size,  $\|f(C) - f(D)\|_F \leq (1 + \sqrt{2})^2 \|f'\|_{\mathbb{D}} \|C - D\|_F$  for a convex open set  $\mathbb{D}$  that contains  $W(C) \cup W(D)$ . The lemma follows by applying this result to the matrices  $V_m^\dagger AV_m$  and  $H_m + h_{m+1,m} \tilde{y} e_m^T$ . The difference of these two matrices has norm  $\|h_{m+1,m}(y - \tilde{y}) e_m^T\|_F = \|h_{m+1,m}(y - \tilde{y})\|_2$ .  $\square$

As the norm of  $V_m$  can be expected to remain modest in our algorithms, Lemma 4.2 implies that the error when solving the least-squares problem will not get amplified significantly, provided that the derivative  $f'$  behaves nicely and that the numerical ranges of the matrices  $V_m^\dagger AV_m$  and  $H_m + h_{m+1,m} \tilde{y} e_m^T$  do not become too large.



**4.2. Analysis of Algorithm 3.3.**

**THEOREM 4.3.** *Considering the setting of Algorithm 3.3, assume that  $f$  is holomorphic on a domain containing  $W(V_m^\dagger AV_m)$  and  $W(H_m)$ . Then*

$$(4.1) \quad \|f(V_m^\dagger AV_m)e_1 - f(H_m)e_1\|_2 \leq (2 + 2\sqrt{2}) \min_{p \in \Pi_{m-1}} \|f - p\|_{W(V_m^\dagger AV_m) \cup W(H_m)}.$$

*Proof.* The matrices  $V_m^\dagger AV_m$  and  $H_m$  are upper Hessenberg and differ only in their last columns. Therefore, for every polynomial  $p \in \Pi_{m-1}$  it holds that  $p(V_m^\dagger AV_m)e_1 = p(H_m)e_1$ . This allows us to write

$$\begin{aligned} \|f(V_m^\dagger AV_m)e_1 - f(H_m)e_1\|_2 &= \|f(V_m^\dagger AV_m)e_1 - p(V_m^\dagger AV_m)e_1 + p(H_m)e_1 - f(H_m)e_1\|_2 \\ &\leq \|(f - p)(V_m^\dagger AV_m)e_1\|_2 + \|(f - p)(H_m)e_1\|_2 \\ &\leq \|(f - p)(V_m^\dagger AV_m)\|_2 + \|(f - p)(H_m)\|_2 \\ &\leq (1 + \sqrt{2}) \left( \|f - p\|_{W(V_m^\dagger AV_m)} + \|f - p\|_{W(H_m)} \right), \end{aligned}$$

where the third inequality is due to a result by Crouzeix and Palencia [8]. □

An immediate consequence of Theorem 4.3 is that the output of Algorithm 3.3 is exact for all polynomials of degree up to  $m - 1$  (and such an output coincides with (3.2)). Such an exactness property holds for all algorithms considered in this paper, namely Algorithms 3.1, 3.2, and 3.3 (and hence also sFOM).

By the triangle inequality, we have that

$$\begin{aligned} &\|f(A)b - \|b\|_2 V_m f(H_m)e_1\|_2 \\ &\leq \|f(A)b - \|b\|_2 V_m f(V_m^\dagger AV_m)e_1\|_2 + \|b\|_2 \|V_m f(V_m^\dagger AV_m)e_1 - V_m f(H_m)e_1\|_2. \end{aligned}$$

We can now use Theorem 4.1 to bound the first term and Theorem 4.3 for the second term, in addition to the fact that the spectral norm of  $V_m$  is bounded by  $\sqrt{m}$  because  $V_m$  has columns of unit norm.

While Theorem 4.1 ensures that the outputs of Algorithms 3.1 and 3.2 are close to  $f(A)b$  whenever  $f$  is well approximated by a polynomial of small degree on  $W(A)$ , in order to obtain a good a priori bound on the error of Algorithm 3.3 we need a good polynomial approximation on a larger set, that is, the union of  $W(V_m^\dagger AV_m)$  and  $W(H_m)$ . In general, the right-hand side of (4.1) is difficult to control and, in practice, we sometimes observe that Algorithm 3.3 fails to converge or the convergence is slowed down (see, e.g., Figures 6 and 7).

**4.3. Discussion on numerical stability.** Rigorous numerical stability analysis of any of the methods discussed in this work appears to be challenging. The evaluation of the function of a small matrix  $f(H_m)$  is already a nontrivial computation, for which (backward or forward) stability is established only for limited classes of matrices and functions  $f$  (e.g., normal matrices with  $f = \exp$  [17, Chapter 10]). To nevertheless gain some insight into the numerical behavior of Algorithms 3.1 and 3.2, particularly in terms of how the conditioning of  $V_m$  affects the stability, let us *assume* that  $f(H_m)$  is computed in a backward stable fashion, that is, its computed approximation can be written as  $f(H_m + \Delta H)$ , where  $\|\Delta H\|_2 = \mathcal{O}(\epsilon \|H_m\|_2)$  with the machine precision  $\epsilon$ . Then in the computation of  $f(V_m^\dagger AV_m)$  in (3.2), the backward error in computing  $f(V_m^\dagger AV_m)$  becomes  $\epsilon \|V_m^\dagger AV_m\|_2$ . Letting  $V_m = Q_m R_m$  be a QR factorization, we have  $\epsilon \|V_m^\dagger AV_m\|_2 = \|R_m^{-1} Q_m^T A Q_m R_m\|_2 \leq \|R_m^{-1}\|_2 \|Q_m^T A Q_m\|_2 \|R_m\|_2 = \kappa_2(R_m) \|Q_m^T A Q_m\|_2$ . This suggests that an ill-conditioned basis  $V_m$  may negatively

impact the numerical stability of the algorithms (in practice, the approximation is often good unless  $V_m$  is numerically rank-deficient). It also suggests that a well-conditioned  $V_m$  would yield an approximation similar to the one obtained via full Arnoldi orthogonalization (wherein  $V_m = Q_m$ ), which we observe in our experiments. In practice, convergence is sometimes observed even when  $V_m$  is highly ill-conditioned; this has been studied extensively for the Lanczos method [20], and steepest descent (restarted Arnoldi) [1] when  $A$  is symmetric. However, a complete analysis for the nonsymmetric and randomized case is an open problem.

**5. Numerical experiments.** We now test the proposed Algorithms 3.1, 3.2, and 3.3 on a variety of examples; these are indicated as “BG-f(A),” “NT1-f(A),” and “NT2-f(A)” in the legends, respectively (the letters BG and NT are the initials of the authors of the papers [3, 21] from where Algorithms 2.1 and 2.2 are taken). The baseline method is “Arnoldi,” which corresponds to (3.1). For some examples, we also compare with sFOM [16, Algorithm 1], where the basis is obtained by Algorithm 2.2, and with the restarting algorithm from [13]. In all our examples we set the truncation parameter  $k = 2$  in Algorithm 2.2. The sketch matrix  $\Theta$  is an SRHT of size  $2m \times n$ , where  $m$  is the dimension of the Krylov subspace we are considering, unless we indicate otherwise.

All algorithms were implemented and executed in MATLAB R2022a. For the evaluation of  $f$  on the compressed matrices we used MATLAB `expm` and `sqrtnm` functions. We report the runtime and the relative accuracy  $\frac{\|f(A)b - \widehat{f(A)b}\|_2}{\|f(A)b\|_2}$ , where  $\widehat{f(A)b}$  is the computed approximation to  $f(A)b$  using one of the algorithms. The “exact” value of  $f(A)b$  is computed using `expm` or `sqrtnm` when the size of the matrix is smaller than  $10^4$ ; otherwise we approximate the exact value running sufficiently many iterations of the Arnoldi algorithm. For the implementation of Algorithm 2.1 we have used the implementation available at <https://github.com/obalabanov/randKrylov>, which uses a subsampled Walsh–Hadamard transform for sketching. The least-squares problems were solved using the MATLAB `lsqr` function with tolerance parameter  $10^{-6}$  and for Blendenpik we used the code at <https://github.com/haimav/Blendenpik>. The code for sFOM was taken from [https://github.com/MarcelSchweitzer/sketched\\_fAb](https://github.com/MarcelSchweitzer/sketched_fAb). The code for the restarted algorithm was taken from [http://www.guettel.com/funm\\_quad/](http://www.guettel.com/funm_quad/). All numerical experiments in this work have been run on a laptop with an eight-core Intel Core i7-8650U 1.90 GHz CPU and 16 GB of RAM. The code for reproducing the experiments is available at <https://github.com/Alice94/RandMatrixFunctions>.

### 5.1. Behavior of Algorithms 3.1, 3.2, and 3.3.

**5.1.1. Convection-diffusion problem.** As a first example, we consider the matrix  $A$  coming from the discretization of the convection-diffusion problem considered in [7, section 3.1]:

$$L[u] = -(D_1 u_x)_x - (D_2 u_y)_y + Pe \left( \frac{1}{2} (v_1 u_x + v_2 u_y) + \frac{1}{2} \left( (v_1 u)_x + (v_2 u)_y \right) \right),$$

$$D_1(x, y) = \begin{cases} 1000, & (x, y) \in [0.25, 0.75]^2, \\ 1 & \text{otherwise,} \end{cases} \quad D_2(x, y) = \frac{1}{2} D_1(x, y),$$

$$v_1(x, y) = x + y, \quad v_2(x, y) = x - y.$$

We take homogeneous Dirichlet boundary conditions and Peclet number  $Pe = 200$ , and we consider a discretization by finite differences on a mesh of size  $352 \times 352$ . The function  $f$  is the exponential and the starting vector is set to the values of the function

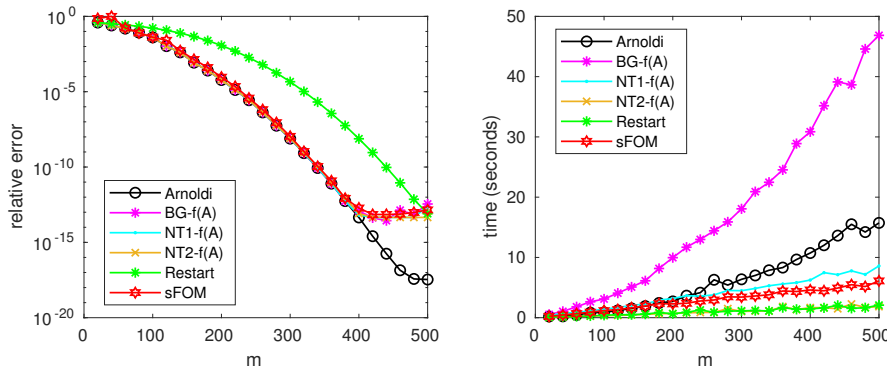


FIG. 1. Exponential of the matrix from the convection-diffusion problem in section 5.1.1.

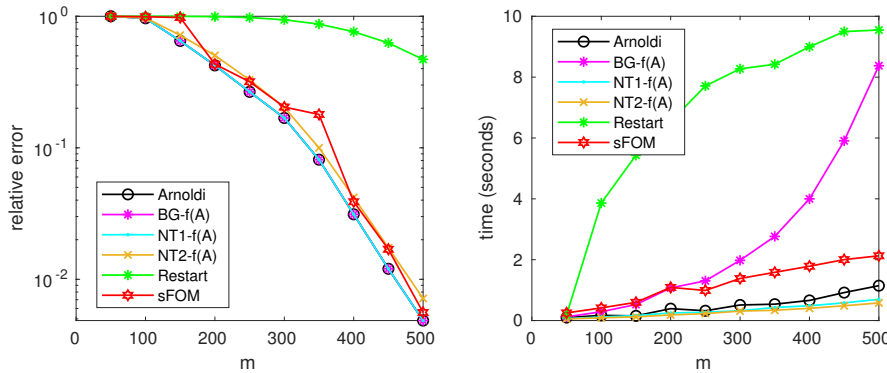


FIG. 2. Computing the exponential of the nonsymmetric Toeplitz matrix from section 5.1.2.

$\sin(\pi x) \sin(\pi y)$  on the finite-difference mesh and then normalized as  $b := b/\|b\|_2$ . The results are reported in Figure 1. We observe that all our proposed methods are stable in this example and they produce approximately the same result. As expected, Algorithm 3.3 is by far the fastest one among our proposed algorithms. Note that the restarted algorithm (with restart length 20) takes approximately the same time for a fixed  $m$  but converges more slowly. The convergence of restarted methods can be often improved, sometimes significantly, by keeping an augmentation space of a few vectors across restart cycles; we refer the reader to [11] for more details.

**5.1.2. Option pricing problem.** Another example is illustrated in Figure 2. We compute the exponential of the nonsymmetric Toeplitz matrix from section 6.3 in [18], size  $n = 2000$ , and we take  $b$  to be a normalized random vector. For a restart length 50, the restarted algorithm exhibits significantly delayed convergence.

**5.1.3. Whitening the basis.** We now consider two examples in which Algorithm 2.2 alone does not give a well-conditioned basis, and therefore whitening (see section 2.1.2) is needed to make the algorithm work. To check the condition number of the sketch  $\Theta V_i$  we use the MATLAB `cond` function and we perform whitening if this exceeds 1000.

We compute the square root of the graph Laplacian of the matrix `p2p_Gnute11a08` from <https://sparse.tamu.edu>, which was considered in [5]. The matrix has size  $n = 6301$ , it is nonsymmetric, and all its eigenvalues have positive real part. We choose  $b$

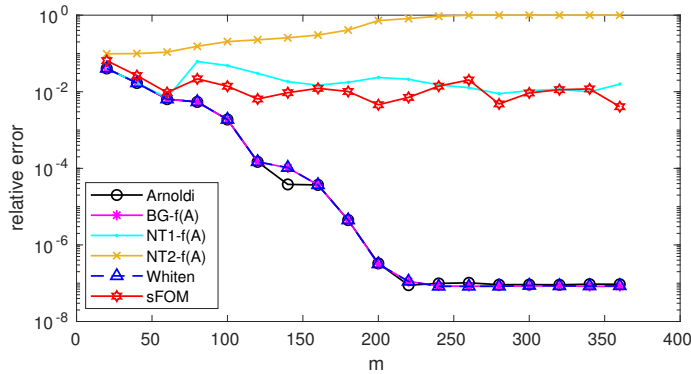


FIG. 3. Square root of the graph Laplacian of the matrix from section 5.1.3. (Color online.)

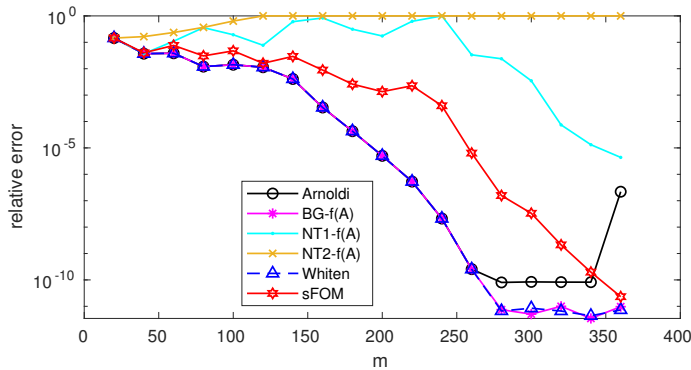


FIG. 4. Sign of the matrix in section 5.1.3.

to be a random unit vector. The results are reported in Figure 3. Around the 60th iteration of Algorithm 2.2, the quality of the basis (in terms of conditioning) starts to deteriorate quickly, and indeed the solid cyan line does not converge. However, when whitening is performed, the convergence is restored.

As another example, we consider the matrix `bfw782a` from <https://math.nist.gov/MatrixMarket/data/NEP/bfwave/bfw782a.html>, a random unit vector  $b$ , and the sign function. To make computations more stable, to apply the sign function to a vector we use the formula  $f(A)b = (A^2)^{-1/2}(Ab)$  [17, Chapter 5]. In this case, the basis obtained from Algorithm 2.2 is badly conditioned and whitening is needed. The results are reported in Figure 4.

**5.1.4. Timing of Algorithm 3.1 versus Arnoldi.** We observed that Algorithm 3.1 is slower than Arnoldi for the example in Figure 1. As discussed in [3], the benefits of Algorithm 2.1 compared to Arnoldi are only visible for moderately large values of  $m$ , the dimension of the Krylov subspace. In Algorithm 3.1, the construction of the basis needs to be combined with an additional step—the solution of the least-squares problem. In Figure 5 we compute the inverse square root of a matrix  $A$  which is the three-dimensional Laplacian matrix of the form

$$T_N \otimes I_N \otimes I_N + I_N \otimes T_N \otimes I_N + I_N \otimes I_N \otimes T_N,$$

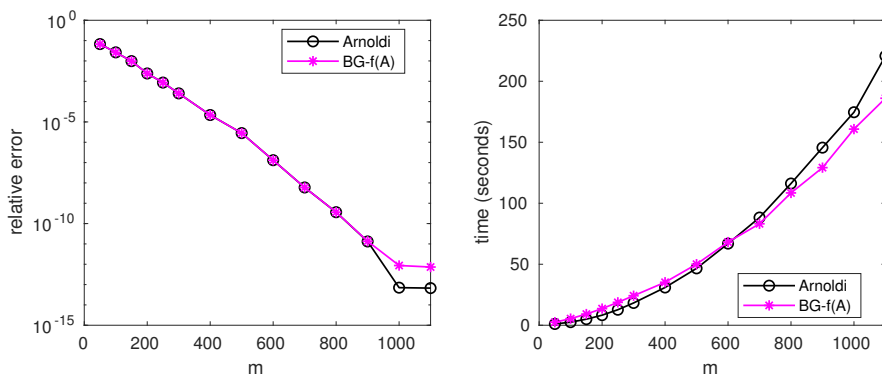


FIG. 5. Computing the inverse square root of the modified three-dimensional Laplacian from section 5.1.4.

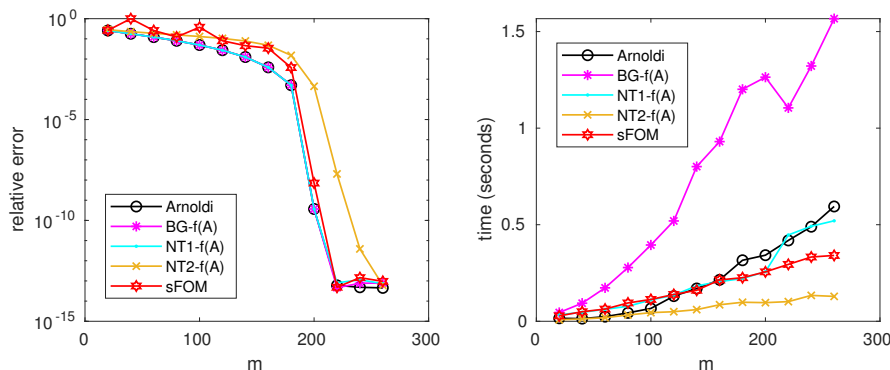


FIG. 6. Convection diffusion example from section 5.2, size  $n = 10^4$ ,  $f$  is the inverse square root.

where  $T_N = \text{tridiag}(-1, 2, 1)$  and  $I_N$  is the identity matrix with size  $N = 80$  (therefore  $n = 512000$ ), perturbed in such a way that the entries on the 10th superdiagonal are 0.125. The vector  $b$  is a randomly chosen unit vector. The sketch matrix  $\Theta$  is of size  $[1.05m] \times n$  in this example. We observe that, for large enough values of  $m$ , Algorithm 3.1 is indeed faster than Arnoldi.

**5.2. Comparison with [16].** We also include a comparison of our algorithms with sFOM [16] on the examples from [16, sections 5.1 and 5.2]. In Figure 6 we consider the inverse square root of a matrix coming from a convection-diffusion problem (different from the one in Figure 1) and in Figure 7 we consider the exponential of the wiki-vote matrix, of size  $n = 8297$ , from the SNAP collection <https://sparse.tamu.edu/SNAP/wiki-Vote>. Observe that, in Figure 7, Algorithm 3.3 stagnates until  $m = 25$ , even though the basis obtained from Algorithm 2.2 is numerically full rank, perhaps reflecting the discussion on stability in section 4.3.

**6. Conclusion.** In this work we proposed two ways of speeding up Krylov subspace methods for computing  $f(A)b$  for a nonsymmetric matrix  $A$  and general function  $f$ . For well-behaved problems, where the condition number of the constructed Krylov subspace basis does not deteriorate, it is best to run Algorithm 3.3 until convergence. Once this condition number gets very large, our experiments indicate that it is beneficial to perform whitening and then switch to the more expensive Algorithm 3.1. The algorithm sFOM proposed in [16] offers a compromise between the different methods,

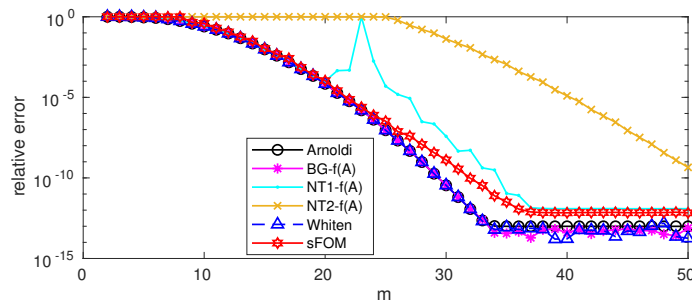


FIG. 7. Exponential of the wiki-vote matrix considered in section 5.2 and in [16, section 5.2].

in being more stable (but slower) than Algorithm 3.3 and faster (but less stable) than Algorithm 3.1.

**Acknowledgments.** The authors thank Laura Grigori, Stefan Güttel, and Marcel Schweitzer for helpful discussions on this work. AC thanks the Mathematical Institute at the University of Oxford for hosting her during a research visit that initiated this work.

#### REFERENCES

- [1] M. AFANASJEW, M. EIERMANN, O. G. ERNST, AND S. GÜTTEL, *A generalization of the steepest descent method for matrix functions*, Electron. Trans. Numer. Anal., 28 (2008), pp. 206–222.
- [2] H. AVRON, P. MAYMOUNKOV, AND S. TOLEDO, *Blendenpik: Supercharging LAPACK’s least-squares solver*, SIAM J. Sci. Comput., 32 (2010), pp. 1217–1236, <https://doi.org/10.1137/090767911>.
- [3] O. BALABANOV AND L. GRIGORI, *Randomized Gram–Schmidt process with application to GMRES*, SIAM J. Sci. Comput., 44 (2022), pp. A1450–A1474, <https://doi.org/10.1137/20M138870X>.
- [4] B. BECKERMANN, A. CORTINOVIS, D. KRESSNER, AND M. SCHWEITZER, *Low-rank updates of matrix functions II: Rational Krylov methods*, SIAM J. Numer. Anal., 59 (2021), pp. 1325–1347, <https://doi.org/10.1137/20M1362553>.
- [5] M. BENZI AND P. BOITO, *Matrix functions in network analysis*, GAMM-Mitt., 43 (2020), e202000012.
- [6] M. BENZI, P. BOITO, AND N. RAZOUK, *Decay properties of spectral projectors with applications to electronic structure*, SIAM Rev., 55 (2013), pp. 3–64, <https://doi.org/10.1137/100814019>.
- [7] M. A. BOTCHEV AND L. A. KNIZHNERMAN, *ART: Adaptive residual-time restarting for Krylov subspace matrix exponential evaluations*, J. Comput. Appl. Math., 364 (2020), 112311, <https://doi.org/10.1016/j.cam.2019.06.027>.
- [8] M. CROUZEIX AND C. PALENCIA, *The numerical range is a  $(1 + \sqrt{2})$ -spectral set*, SIAM J. Matrix Anal. Appl., 38 (2017), pp. 649–655, <https://doi.org/10.1137/17M1116672>.
- [9] P. I. DAVIES AND N. J. HIGHAM, *A Schur–Parlett algorithm for computing matrix functions*, SIAM J. Matrix Anal. Appl., 25 (2003), pp. 464–485, <https://doi.org/10.1137/S0895479802410815>.
- [10] M. EIERMANN AND O. G. ERNST, *A restarted Krylov subspace method for the evaluation of matrix functions*, SIAM J. Numer. Anal., 44 (2006), pp. 2481–2504, <https://doi.org/10.1137/050633846>.
- [11] M. EIERMANN, O. G. ERNST, AND S. GÜTTEL, *Deflated restarting for matrix functions*, SIAM J. Matrix Anal. Appl., 32 (2011), pp. 621–641, <https://doi.org/10.1137/090774665>.
- [12] E. ESTRADA AND D. J. HIGHAM, *Network properties revealed through matrix functions*, SIAM Rev., 52 (2010), pp. 696–714, <https://doi.org/10.1137/090761070>.
- [13] A. FROMMER, S. GÜTTEL, AND M. SCHWEITZER, *Efficient and stable Arnoldi restarts for matrix functions based on quadrature*, SIAM J. Matrix Anal. Appl., 35 (2014), pp. 661–683, <https://doi.org/10.1137/13093491X>.

- [14] S. GÜTTEL, *Rational Krylov approximation of matrix functions: Numerical methods and optimal pole selection*, GAMM-Mitt., 36 (2013), pp. 8–31, <https://doi.org/10.1002/gamm.201310002>.
- [15] S. GÜTTEL, D. KRESSNER, AND K. LUND, *Limited-memory polynomial methods for large-scale matrix functions*, GAMM-Mitt., 43 (2020), e202000019.
- [16] S. GÜTTEL AND M. SCHWEITZER, *Randomized Sketching for Krylov Approximations of Large-Scale Matrix Functions*, preprint, arXiv:2208.11447, 2022.
- [17] N. J. HIGHAM, *Functions of Matrices*, SIAM, Philadelphia, 2008, <https://doi.org/10.1137/1.9780898717778>.
- [18] D. KRESSNER AND R. LUCE, *Fast computation of the matrix exponential for a Toeplitz matrix*, SIAM J. Matrix Anal. Appl., 39 (2018), pp. 23–47, <https://doi.org/10.1137/16M1083633>.
- [19] P.-G. MARTINSSON AND J. A. TROPP, *Randomized numerical linear algebra: Foundations and algorithms*, Acta Numer., 29 (2020), pp. 403–572.
- [20] C. MUSCO, C. MUSCO, AND A. SIDFORD, *Stability of the Lanczos method for matrix function approximation*, in Proceedings of the 29th Annual ACM-SIAM Symposium on Discrete Algorithms, SIAM, Philadelphia, 2018, pp. 1605–1624.
- [21] Y. NAKATSUKASA AND J. A. TROPP, *Fast & Accurate Randomized Algorithms for Linear Systems and Eigenvalue Problems*, preprint, arXiv:2111.00113, 2021.
- [22] C. C. PAIGE AND M. A. SAUNDERS, *LSQR: An algorithm for sparse linear equations and sparse least squares*, ACM Trans. Math. Software, 8 (1982), pp. 43–71, <https://doi.org/10.1145/355984.355989>.
- [23] Y. SAAD, *Numerical Methods for Large Eigenvalue Problems*, Wiley, New York, 1992.
- [24] Y. SAAD, *Iterative Methods for Sparse Linear Systems*, SIAM, Philadelphia, 2003, <https://doi.org/10.1137/1.9780898718003>.
- [25] J. A. TROPP, *Improved analysis of the subsampled randomized Hadamard transform*, Adv. Adapt. Data Anal., 3 (2011), pp. 115–126, <https://doi.org/10.1142/S1793536911000787>.
- [26] R. ČIEGIS, V. STARIKOVIČIUS, S. MARGENOV, AND R. KRIAUIENĖ, *A comparison of accuracy and efficiency of parallel solvers for fractional power diffusion problems*, in Parallel Processing and Applied Mathematics. Part I, Lecture Notes in Comput. Sci. 10777, Springer, Cham, 2018, pp. 79–89, <https://doi.org/10.1007/978-3-319-78024-5>.