Bergische Universität Wuppertal

Fachbereich Mathematik und Naturwissenschaften

Institute of Mathematical Modelling, Analysis and Computational Mathematics
(IMACM)

Andreas Frommer, Kathryn Lund and Daniel B. Szyld

Block Krylov subspace methods
for computing functions of matrices
applied to multiple vectors

http://www.imacm.uni-wuppertal.de

BLOCK KRYLOV SUBSPACE METHODS FOR COMPUTING FUNCTIONS OF MATRICES APPLIED TO MULTIPLE VECTORS

ANDREAS FROMMER[*], KATHRYN LUND[†,†], AND DANIEL B. SZYLD[‡,‡]

Abstract. A variety of block Krylov subspace methods have been successfully developed for linear systems and matrix equations. The application of block Krylov methods to computing matrix functions is, however, less established, despite the growing prevalence of matrix functions in scientific computing. Of particular importance is the evaluation of a matrix function on not just one but multiple vectors. The main contribution of this paper is a class of efficient block Krylov subspace methods tailored precisely to this task. With the full orthogonalization method (FOM) for linear systems forming the backbone of our theory, the resulting methods are referred to as B(FOM)^2: block FOM for functions of matrices.

Many other important results are obtained in the process of developing these new methods. Matrix-valued inner products are used to construct a general framework for block Krylov subspaces that encompasses already established results in the literature. Convergence bounds for B(FOM)^2 are proven for Stieltjes functions applied to a class of matrices which are self-adjoint and positive definite with respect to the matrix-valued inner product. A detailed algorithm for B(FOM)^2 with restarts is developed. Its efficiency is based on a recursive expression for the error, which is also used to update the solution. Numerical experiments demonstrate the power and versatility of this new class of methods for a variety of matrix-valued inner products, functions, and matrices.

Key words. matrix functions, restarted Krylov subspace methods, block Krylov subspace methods, global methods

AMS subject classifications. 65F60, 65F50, 65F10, 65F30

1. Introduction. Many applications in scientific computing require the efficient computation of $f(A)\boldsymbol{B}$, where f is a scalar function defined on an $n \times n$ matrix A, and $\boldsymbol{B}$ is a *block vector* in $\mathbb{C}^{n \times s}$. Regarding $\boldsymbol{B}$ as a collection of columns $\boldsymbol{b}_i$,

$$\boldsymbol{B} := [\boldsymbol{b}_1|...|\boldsymbol{b}_s],$$

one might consider applying methods for a single vector, such as those described in [12, 27] or the newly proposed restarted Arnoldi methods [24, 25, 26], to each problem $f(A)\boldsymbol{b}_i$. It is well known for linear systems, however, that block Krylov approaches treating all columns $\boldsymbol{b}_i$ at once can be computationally advantageous; see, e.g., [4, 33, 45, 49, 51, 58, 59, 60, 62]. It is therefore reasonable to consider block Krylov methods for computing $f(A)\boldsymbol{B}$.

Others have considered block Krylov methods for $f(A)\boldsymbol{B}$ before. Lopez and Simoncini developed a block Krylov method for $\exp(A)\boldsymbol{B}$ so that so-called geometric properties of $\boldsymbol{B}$ are preserved, but did not undertake a convergence analysis [43]. Benner, Kürschner, and Saak applied Krlyov and integral-based methods to compute $\log(A)\boldsymbol{B}$, but did not develop block-based theory in any detail [6]. Al-Mohy and Higham proposed a direct method for computing $\exp(A)\boldsymbol{B}$, based on the scaling and squaring method and a truncated Taylor series approximation [2]. This method is indeed dominated by matrix-block-vector multiplications $A\boldsymbol{B}$, which can make it superior to Krylov methods.

However, the two Krylov methods the authors of [2] compare it with do not feature blocks, unlike the ones we propose here.

Our development draws from two main sources: Krylov methods for matrix functions, and the block full orthogonalization method (BFOM) for linear systems, which reduces to block conjugate gradients (BCG) when the matrix considered is Hermitian positive definite (HPD).

1.1. Krylov methods for $f(A)\boldsymbol{b}$. There is an expansive body of literature devoted to methods for computing $f(A)\boldsymbol{b}$, for f more general than the exponential or logarithm, and $\boldsymbol{b}$ a single vector. We focus on methods that approximate $f(A)\boldsymbol{b}$ by a Krylov subspace, such as those in [1, 7, 23, 24, 25, 26, 39, 53]. These methods have been referred to as Arnoldi or Lanczos methods for matrix functions, depending on whether the matrix A is Hermitian. For the sake of the framework developed herein, we propose a new nomenclature. We refer to all such methods as full orthogonalization methods (FOM) for functions of matrices, or $(\text{FOM})^2$ for short, due to the inherent connections between these methods and FOM theory, which we recapitulate in section 3. In particular, we generalize [24, 25], which use quadrature to compute stable restarted approximations to $f(A)\boldsymbol{b}$, where f is a Stieltjes function.

1.2. Block Krylov methods. Block Krylov subspace methods for computing $A^{-1}\boldsymbol{B}$ comprise an even larger body of literature, inadvertently causing some ambiguity as to what one means by a "block" Krlyov method. We propose the following three categories to describe the variety of methods found in the literature:

Classical block methods. These are methods related to Dianne O'Leary's 1980 seminal paper on block cojugate gradients [49]. BCG has been explored thoroughly in the literature [4, 10, 17, 47, 48, 52], along with other block methods, such as BGMRES [33, 34, 45, 57, 58, 59, 60]. A survey of classical block Krylov methods is given in [55].

Global block methods. Global GMRES and global FOM were first introduced in 1999 in [40] for matrix equations. Additional global methods can be found in [3, 9, 22, 36, 51, 66].

Loop-interchange block methods. These methods are not as historically established as the classical and global methods. Rather, they have been proposed as an implementation alternative to the non-block approach in which the systems are solved serially for one column after another. While it is likely that such methods have been considered before, they are first formally named and developed in [51].

We denote the FOM variant of each of the above classes as ClBFOM, GlBFOM, and LiBFOM, respectively. A detailed description clarifying what precisely differentiates all three methods will be developed later in this paper.

1.3. Advantages of blocks. There is no difference between the *matrix-block product $A\boldsymbol{V}$* and the concatenation of *matrix-vector products $[A\boldsymbol{v}_1|\ldots|A\boldsymbol{v}_s]$*; indeed, $A\boldsymbol{V} = [A\boldsymbol{v}_1|\ldots|A\boldsymbol{v}_s]$. However, from the standpoint of computational efficiency, formulating an algorithm in terms of matrix-block products leads to an improved process, for a number of reasons described in, e.g., [5, 16, 18, 38]. For one, matrix-block products only require that A be accessed once, versus s times if each column $A\boldsymbol{v}_i$ is computed independently. Reducing the number of times A must be accessed is advantageous when computer architecture (i.e., machine-specific properties of the CPU, GPU, or other hardware accelerators) renders memory access costly. The advantage increases when A is not explicitly stored or must be generated anew each time it is accessed. In a parallel environment, block methods present additional advantages since they reduce communication among processors.

Building a Krylov subspace based on block vectors instead of column vectors also "enriches" the theoretical Krylov subspace from which approximations are drawn. Classical block methods,

for example, use information from all columns of $\boldsymbol{B}$ to generate approximations to each column of $\boldsymbol{X}$, leading to potentially more accurate approximations per iteration and consequently faster convergence. Methods such as multi-preconditioned GMRES [30], enlarged Krylov subspace conjugate gradients [31], and augmented Krylov subspace methods [61, section 9] apply this principle even more generally to speed up convergence. We do not, however, explore such techniques further in the present work.

1.4. Notation. Let $\otimes$ denote the Kronecker product between two matrices. With I_d denoting the identity matrix of size $d \times d$, we define the standard unit vector $\widehat{\boldsymbol{e}}_k^d \in \mathbb{C}^d$ as the kth column of I_d. The standard unit block vector $\widehat{\boldsymbol{E}}_k \in \mathbb{C}^{ms \times s}$ is defined as $\widehat{\boldsymbol{E}}_k := \widehat{\boldsymbol{e}}_k^m \otimes I_s$. We do not distinguish notationally between the matrix $A \in \mathbb{C}^{n \times n}$ as an operator on $\mathbb{C}^n$, $\boldsymbol{v} \to A\boldsymbol{v}$ and as an operator on $\mathbb{C}^{n \times s}$, $\boldsymbol{V} = [\boldsymbol{v}_1 \mid \cdots \mid \boldsymbol{v}_s] \to A\boldsymbol{V} = [A\boldsymbol{v}_1 \mid \cdots \mid A\boldsymbol{v}_s]$.

1.5. Outline. In section 2, we introduce the concept of a matrix-valued inner product $\langle\!\langle \cdot, \cdot \rangle\!\rangle_{\mathbb{S}}$ for block vectors with values in a *-algebra $\mathbb{S} \subset \mathbb{C}^{s \times s}$. The *- algebra $\mathbb{S}$ together with $\langle\!\langle \cdot, \cdot \rangle\!\rangle_{\mathbb{S}}$ are fundamental in the sense that they yield different kinds of block Krylov subspaces and of block orthogonality. We develop a general block Arnoldi process based on $\langle\!\langle \cdot, \cdot \rangle\!\rangle_{\mathbb{S}}$ encompassing the three classes of block approaches mentioned in section 1.2. In section 3, we show how each choice of $\mathbb{S}$ together with $\langle\!\langle \cdot, \cdot \rangle\!\rangle_{\mathbb{S}}$ determines a specific BFOM for linear equations, and we derive a "true" scalar-valued inner product from $\langle\!\langle \cdot, \cdot \rangle\!\rangle_{\mathbb{S}}$, which allows for convergence analysis of BFOM for matrices self-adjoint and positive definite with respect to $\langle\!\langle \cdot, \cdot \rangle\!\rangle_{\mathbb{S}}$. Section 4 then details our main contributions. We expound the BFOM for functions of matrices (B(FOM)^2), which arise for different $\mathbb{S}$ and $\langle\!\langle \cdot, \cdot \rangle\!\rangle_{\mathbb{S}}$, show how they can be stably restarted, and provide a convergence proof for an important class of functions and matrices. In section 5 we illustrate the performance of the classical, global, and loop-interchange versions of restarted B(FOM)^2 in comparison to the non-block approach with some numerical experiments.

2. Block Krylov subspaces. The use of matrix-valued functions in the analysis of block Krylov methods is not new; see, e.g., the right bilinear form of [21], the $\diamond$ product of [9], or the block inner product of [28]. Neither is the notion of a matrix-valued inner product novel; see, e.g., literature on matrix-valued orthogonal polynomials [11, 32].

Let $\mathbb{S}$ be a *-subalgebra of $\mathbb{C}^{s \times s}$ with identity, i.e., $\mathbb{S}$ is a vector subspace containing I_s that is closed under matrix multiplication and conjugate transposition. $\mathbb{S}$ thus also contains the inverses C^{-1} of all its nonsingular elements $C \in \mathbb{C}^{s \times s}$, since C^{-1} can be expressed as a polynomial in C. The following generalizes the notion of an inner product to a bilinear operator mapping $\mathbb{C}^{n \times s} \times \mathbb{C}^{n \times s}$ to $\mathbb{S}$ and serves as the foundation for our framework.

DEFINITION 2.1. *A mapping* $\langle\!\langle \cdot, \cdot \rangle\!\rangle_{\mathbb{S}}$ *from* $\mathbb{C}^{n \times s} \times \mathbb{C}^{n \times s}$ *to* $\mathbb{S}$ *is called a* block inner product *onto* $\mathbb{S}$ *if it satisfies the following conditions for all* $\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z} \in \mathbb{C}^{n \times s}$ *and* $C \in \mathbb{S}$:

(i) $\mathbb{S}$*-linearity:* $\langle\!\langle \boldsymbol{X}, \boldsymbol{Y}C \rangle\!\rangle_{\mathbb{S}} = \langle\!\langle \boldsymbol{X}, \boldsymbol{Y} \rangle\!\rangle_{\mathbb{S}}C$ *and* $\langle\!\langle \boldsymbol{X} + \boldsymbol{Y}, \boldsymbol{Z} \rangle\!\rangle_{\mathbb{S}} = \langle\!\langle \boldsymbol{X}, \boldsymbol{Z} \rangle\!\rangle_{\mathbb{S}} + \langle\!\langle \boldsymbol{Y}, \boldsymbol{Z} \rangle\!\rangle_{\mathbb{S}}$,

(ii) *symmetry:* $\langle\!\langle \boldsymbol{X}, \boldsymbol{Y} \rangle\!\rangle_{\mathbb{S}} = \langle\!\langle \boldsymbol{Y}, \boldsymbol{X} \rangle\!\rangle_{\mathbb{S}}^*$,

(iii) *definiteness:* $\langle\!\langle \boldsymbol{X}, \boldsymbol{X} \rangle\!\rangle_{\mathbb{S}}$ *is positive definite if* $\boldsymbol{X}$ *has full rank, and* $\langle\!\langle \boldsymbol{X}, \boldsymbol{X} \rangle\!\rangle_{\mathbb{S}} = 0_s$ *if and only if* $\boldsymbol{X} = \boldsymbol{0}$.

Note that by condition (ii) the matrix $\langle\!\langle \boldsymbol{X}, \boldsymbol{X} \rangle\!\rangle_{\mathbb{S}}$ is always Hermitian, and conditions (i) and (ii) together imply that $\langle\!\langle \boldsymbol{X} + \boldsymbol{Y}C, \boldsymbol{Z} \rangle\!\rangle_{\mathbb{S}} = \langle\!\langle \boldsymbol{X}, \boldsymbol{Z} \rangle\!\rangle_{\mathbb{S}} + C^*\langle\!\langle \boldsymbol{Y}, \boldsymbol{Z} \rangle\!\rangle_{\mathbb{S}}$. Also, since $\langle\!\langle \cdot, \cdot \rangle\!\rangle_{\mathbb{S}}$ is continuous, (iii) implies that $\langle\!\langle \boldsymbol{X}, \boldsymbol{X} \rangle\!\rangle_{\mathbb{S}}$ is positive semidefinite when $\boldsymbol{X}$ is rank-deficient. The following definition is related to a generalization of the notion of norm.

DEFINITION 2.2. *A mapping* N *which maps all* $\boldsymbol{X} \in \mathbb{C}^{n \times s}$ *with full rank on a matrix* $N(\boldsymbol{X}) \in \mathbb{S}$

is called a scaling quotient *if for all such* $\boldsymbol{X}$ *there exists* $\boldsymbol{Y} \in \mathbb{C}^{n \times s}$ *such that* $\boldsymbol{X} = \boldsymbol{Y}N(\boldsymbol{X})$ *and* $\langle\!\langle \boldsymbol{Y}, \boldsymbol{Y} \rangle\!\rangle_{\mathbb{S}} = I_s$.

Clearly, $N(\boldsymbol{X})$ is nonsingular, since for $\boldsymbol{X}$ full rank the matrix

$$\langle\!\langle \boldsymbol{X}, \boldsymbol{X} \rangle\!\rangle_{\mathbb{S}} = N(\boldsymbol{X})^* \langle\!\langle \boldsymbol{Y}, \boldsymbol{Y} \rangle\!\rangle_{\mathbb{S}} N(\boldsymbol{X}) = N(\boldsymbol{X})^* N(\boldsymbol{X})$$

is positive definite.[1] For arbitrary $\mathbb{S}$, the Hermitian matrix $\langle\!\langle \boldsymbol{X}, \boldsymbol{X} \rangle\!\rangle_{\mathbb{S}}^{1/2}$ lies in $\mathbb{S}$, since it can be represented as a polynomial in $\langle\!\langle \boldsymbol{X}, \boldsymbol{X} \rangle\!\rangle_{\mathbb{S}}$. It is thus a scaling quotient with $\boldsymbol{Y} = \boldsymbol{X}\langle\!\langle \boldsymbol{X}, \boldsymbol{X} \rangle\!\rangle_{\mathbb{S}}^{-1/2}$. Provided it lies in $\mathbb{S}$, the Cholesky factor of $\langle\!\langle \boldsymbol{X}, \boldsymbol{X} \rangle\!\rangle_{\mathbb{S}}$ is a scaling quotient as well. We call a pair of block vectors $\boldsymbol{X}$ and $\boldsymbol{Y}$ $\langle\!\langle \cdot, \cdot \rangle\!\rangle_{\mathbb{S}}$-*orthogonal*, if $\langle\!\langle \boldsymbol{X}, \boldsymbol{Y} \rangle\!\rangle_{\mathbb{S}} = 0_s$, and we call a block vector $\boldsymbol{X}$ $\langle\!\langle \cdot, \cdot \rangle\!\rangle_{\mathbb{S}}$-*normalized* if $\langle\!\langle \boldsymbol{X}, \boldsymbol{X} \rangle\!\rangle_{\mathbb{S}} = I_s$. Combining the two notions, a set of block vectors $\{\boldsymbol{X}_1, \ldots, \boldsymbol{X}_m\}$ is called $\langle\!\langle \cdot, \cdot \rangle\!\rangle_{\mathbb{S}}$-*orthonormal* when $\langle\!\langle \boldsymbol{X}_i, \boldsymbol{X}_j \rangle\!\rangle_{\mathbb{S}} = \delta_{ij} I_s$, where δ_{ij} is the Kronecker delta. Clearly, if $\boldsymbol{X}$ has full rank, $\boldsymbol{X}N(\boldsymbol{X})^{-1}$ is $\langle\!\langle \cdot, \cdot \rangle\!\rangle_{\mathbb{S}}$-normalized.

The following three combinations of $\mathbb{S}$, $\langle\!\langle \cdot, \cdot \rangle\!\rangle_{\mathbb{S}}$ and N, whose block inner products were first deduced in [22], satisfy the above definitions.

EXAMPLE 2.3 (Classical). $\mathbb{S}^{\mathrm{Cl}} = \mathbb{C}^{s \times s}$ and $\langle\!\langle \boldsymbol{X}, \boldsymbol{Y} \rangle\!\rangle_{\mathbb{S}}^{\mathrm{Cl}} = \boldsymbol{X}^* \boldsymbol{Y}$. The scaling quotient $N^{\mathrm{Cl}}(\boldsymbol{X})$ is the factor R from the economical QR factorization of $\boldsymbol{X}$, i.e., $\boldsymbol{X} = \boldsymbol{Q}R$ with $\boldsymbol{Q} \in \mathbb{C}^{n \times s}$ and $R \in \mathbb{C}^{s \times s}$. Equivalently, $R = L^*$, with L denoting the Cholesky factor of $\langle\!\langle \boldsymbol{X}, \boldsymbol{X} \rangle\!\rangle_{\mathbb{S}}^{\mathrm{Cl}} = \boldsymbol{X}^* \boldsymbol{X}$.

EXAMPLE 2.4 (Global). $\mathbb{S}^{\mathrm{Gl}}$ is the set of scalar multiples of I_s, and $\langle\!\langle \boldsymbol{X}, \boldsymbol{Y} \rangle\!\rangle_{\mathbb{S}}^{\mathrm{Gl}} = \mathrm{trace}(\boldsymbol{X}^* \boldsymbol{Y})I_s$. The scaling quotient is given as $N^{\mathrm{Gl}}(\boldsymbol{X}) = \|\boldsymbol{X}\|_{\mathrm{F}} I_s$.

EXAMPLE 2.5 (Loop-interchange). $\mathbb{S}^{\mathrm{Li}}$ is the set of diagonal matrices, and $\langle\!\langle \boldsymbol{X}, \boldsymbol{Y} \rangle\!\rangle_{\mathbb{S}}^{\mathrm{Li}} = \mathrm{diag}(\boldsymbol{X}^* \boldsymbol{Y})$, where diag is the operator that sets to zero all the off-diagonal entries of a given square matrix while preserving its diagonal. For $\boldsymbol{X} = [\boldsymbol{x}_1 | \ldots | \boldsymbol{x}_s]$ the scaling quotient of $\boldsymbol{X}$ is given by

$$N^{\mathrm{Li}}(\boldsymbol{X}) = \begin{bmatrix} \|\boldsymbol{x}_1\|_2 & & \\ & \ddots & \\ & & \|\boldsymbol{x}_s\|_2 \end{bmatrix}.$$

Further interesting block inner products exist. One might consider "hybrid" block inner products that combine two of the above three examples. For example, if $s = pq$ we can take $\mathbb{S}^{\mathrm{Hy}}$ as the set of all block diagonal matrices with blocks of size $q \times q$. For $\boldsymbol{X} = [\boldsymbol{X}_1 | \cdots | \boldsymbol{X}_p] \in \mathbb{C}^{n \times s}$ with $\boldsymbol{X}_i \in \mathbb{C}^{n \times q}$, and similarly for $\boldsymbol{Y} \in \mathbb{C}^{n \times s}$, we can then define

$$\langle\!\langle \boldsymbol{X}, \boldsymbol{Y} \rangle\!\rangle_{\mathbb{S}}^{\mathrm{Hy}} = \begin{bmatrix} \boldsymbol{X}_1^* \boldsymbol{Y}_1 & & \\ & \ddots & \\ & & \boldsymbol{X}_p^* \boldsymbol{Y}_p \end{bmatrix}, \tag{2.1}$$

which is a hybrid of $\langle\!\langle \boldsymbol{X}, \boldsymbol{Y} \rangle\!\rangle_{\mathbb{S}}^{\mathrm{Cl}}$ and $\langle\!\langle \boldsymbol{X}, \boldsymbol{Y} \rangle\!\rangle_{\mathbb{S}}^{\mathrm{Li}}$, in some sense.

---

[1] This equality also explains why we do not extend the concept of a scaling quotient to rank deficient $\boldsymbol{X}$. Then, at least one of the two, $\boldsymbol{Y}$ or $N(\boldsymbol{X})$ are necessarily rank-deficient, and it seems impossible to guarantee that $\boldsymbol{X} = \boldsymbol{Y}N(\boldsymbol{X})$ can be fulfilled for general $\langle\!\langle \cdot, \cdot \rangle\!\rangle_{\mathbb{S}}$.

For our work, the following natural extension of self-adjointness to blocks turns out to be the most appropriate.

DEFINITION 2.6. *A matrix $A \in \mathbb{C}^{n \times n}$ is $\langle\!\langle \cdot, \cdot \rangle\!\rangle_{\mathbb{S}}$-self-adjoint if for all $\boldsymbol{X}, \boldsymbol{Y} \in \mathbb{C}^{n \times s}$,*

$$\langle\!\langle A\boldsymbol{X}, \boldsymbol{Y} \rangle\!\rangle_{\mathbb{S}} = \langle\!\langle \boldsymbol{X}, A\boldsymbol{Y} \rangle\!\rangle_{\mathbb{S}}.$$

For the block inner products $\langle\!\langle \cdot, \cdot \rangle\!\rangle_{\mathbb{S}}$ from Examples 2.3-2.5, A being $\langle\!\langle \cdot, \cdot \rangle\!\rangle_{\mathbb{S}}$-self-adjoint reduces to A being Hermitian in the usual sense, i.e., $A = A^*$.

DEFINITION 2.7. *Given a set of block vectors $\{\boldsymbol{X}_k\}_{k=1}^m \subset \mathbb{C}^{n \times s}$, the $\mathbb{S}$-span of $\{\boldsymbol{X}_k\}_{k=1}^m$ is defined as*

$$\mathrm{span}^{\mathbb{S}}\{\boldsymbol{X}_1, \ldots, \boldsymbol{X}_m\} := \left\{ \sum_{k=1}^m \boldsymbol{X}_k C_k : C_k \in \mathbb{S} \right\}. \tag{2.2}$$

Note that $\mathrm{span}^{\mathbb{S}}\{\boldsymbol{X}_1, \ldots, \boldsymbol{X}_m\}$ is indeed a subspace of $\mathbb{C}^{n \times s}$. We can then define the *mth block Krylov subspace with respect to A, $\boldsymbol{B}$, and $\mathbb{S}$* as

$$\mathscr{K}_m^{\mathbb{S}}(A, \boldsymbol{B}) := \mathrm{span}^{\mathbb{S}}\{\boldsymbol{B}, A\boldsymbol{B}, \ldots, A^{m-1}\boldsymbol{B}\}. \tag{2.3}$$

To illustrate how the choice of $\mathbb{S}$ affects the structure of $\mathscr{K}_m^{\mathbb{S}}(A, \boldsymbol{B})$, we look at the Krylov subspaces associated to Examples 2.3-2.5. The classical block inner product gives rise to what is historically known as the block Krylov space; see [33, 34] for a more detailed description of this space, along with a useful notion of block grade. On the other hand, the global and loop-interchange block inner products give rise to special subspaces of $\mathscr{K}_m^{\mathrm{Cl}}(A, \boldsymbol{B})$:

$$\mathscr{K}_m^{\mathrm{Cl}}(A, \boldsymbol{B}) = \left\{ \sum_{k=0}^{m-1} A^k \boldsymbol{B} C_k : C_k \in \mathbb{C}^{s \times s} \right\};$$

$$\mathscr{K}_m^{\mathrm{Gl}}(A, \boldsymbol{B}) = \mathrm{span}\{\boldsymbol{B}, A\boldsymbol{B}, \ldots, A^{m-1}\boldsymbol{B}\} = \left\{ \sum_{k=0}^{m-1} A^k \boldsymbol{B} c_k : c_k \in \mathbb{C} \right\},$$

$$\mathscr{K}_m^{\mathrm{Li}}(A, \boldsymbol{B}) = \mathscr{K}_m(A, \boldsymbol{b}_1) \times \cdots \times \mathscr{K}_m(A, \boldsymbol{b}_s) = \left\{ \sum_{k=0}^{m-1} A^k \boldsymbol{B} D_k : D_k \in \mathbb{C}^{s \times s} \text{ is diagonal} \right\},$$

$$\text{where } \mathscr{K}_m(A, \boldsymbol{b}_i) := \mathrm{span}\{\boldsymbol{b}_i, A\boldsymbol{b}_i, \ldots, A^{m-1}\boldsymbol{b}_i\} \subset \mathbb{C}^n.$$

We can determine $\langle\!\langle \cdot, \cdot \rangle\!\rangle_{\mathbb{S}}$-orthonormal bases of these spaces with a block Arnoldi-type process, detailed in Algorithm 1. As in the non-block case, the process simplifies to a Lanczos-type process with a three-term recurrence if A is $\langle\!\langle \cdot, \cdot \rangle\!\rangle_{\mathbb{S}}$-self-adjoint. For this case, one could also derive a coupled two-term recurrence version as is used for BCG in, e.g., [49].

The block Arnoldi process breaks down if $\boldsymbol{W}$ does not have full rank, because then $N(\boldsymbol{W})$ does not exist and $\boldsymbol{W}$ cannot be $\langle\!\langle \cdot, \cdot \rangle\!\rangle_{\mathbb{S}}$-normalized. This case actually represents a lucky breakdown and can be cured by reducing the dimension s accordingly for subsequent iterations. We implement such *deflation* in our numerical experiments, but, for simplicity of presentation, we do not discuss it any further here. For now, we thus assume that Algorithm 1 runs without breakdowns up to iteration m, i.e., that we obtain

---

Algorithm 1: Block Arnoldi and Block Lanczos

   Given: A, $\boldsymbol{B}$, $\mathbb{S}$, $\langle\langle\cdot,\cdot\rangle\rangle_{\mathbb{S}}$, N, m

1 Compute $B = N(\boldsymbol{B})$ and $\boldsymbol{V}_1 = \boldsymbol{B}B^{-1}$

2 **if** A is $\langle\langle\cdot,\cdot\rangle\rangle_{\mathbb{S}}$-*self-adjoint* **then**

3   Set $\boldsymbol{V}_0 = \boldsymbol{0}$, $H_{0,1} = B$

4   **for** $k = 1, \ldots, m$ **do**

5 $\boldsymbol{W} = A\boldsymbol{V}_k - \boldsymbol{V}_{k-1}H_{k-1,k}$

6 $H_{k,k} = \langle\langle\boldsymbol{V}_k, \boldsymbol{W}\rangle\rangle_{\mathbb{S}}$

7 $\boldsymbol{W} = \boldsymbol{W} - \boldsymbol{V}_kH_{k,k}$

8   Compute $H_{k+1,k} = N(\boldsymbol{W})$ and $\boldsymbol{V}_{k+1} = \boldsymbol{W}H_{k+1,k}^{-1}$

9   Set $H_{k,k+1} = H_{k+1,k}^*$

10 **else**

11   **for** $k = 1, \ldots, m$ **do**

12   Compute $\boldsymbol{W} = A\boldsymbol{V}_k$

13   **for** $j = 1, \ldots, k$ **do**

14 $H_{j,k} = \langle\langle\boldsymbol{V}_j, \boldsymbol{W}\rangle\rangle_{\mathbb{S}}$

15 $\boldsymbol{W} = \boldsymbol{W} - \boldsymbol{V}_jH_{j,k}$

16   Compute $H_{k+1,k} = N(\boldsymbol{W})$ and $\boldsymbol{V}_{k+1} = \boldsymbol{W}H_{k+1,k}^{-1}$

17 Return B, $\boldsymbol{\mathcal{V}}_m = [\boldsymbol{V}_1|\ldots|\boldsymbol{V}_m]$, $\mathcal{H}_m = (H_{j,k})_{j,k=1}^m$, $\boldsymbol{V}_{m+1}$, and $H_{m+1,m}$

   (i) a $\langle\langle\cdot,\cdot\rangle\rangle_{\mathbb{S}}$-orthonormal basis $\{\boldsymbol{V}_k\}_{k=1}^{m+1} \subset \mathbb{C}^{n\times s}$, such that each $\boldsymbol{V}_k$ has full rank and $\mathscr{K}_m^{\mathbb{S}}(A, \boldsymbol{B}) = \mathrm{span}^{\mathbb{S}}\{\boldsymbol{V}_k\}_{k=1}^m$,

   (ii) a block upper Hessenberg matrix $\mathcal{H}_m \in \mathbb{S}^{m\times m}$, and

   (iii) $H_{m+1,m} \in \mathbb{S}$,

all satisfying the *block Arnoldi relation*

$$A\boldsymbol{\mathcal{V}}_m = \boldsymbol{\mathcal{V}}_m\mathcal{H}_m + \boldsymbol{V}_{m+1}H_{m+1,m}\widehat{\boldsymbol{E}}_m^*, \tag{2.4}$$

where $\boldsymbol{\mathcal{V}}_m = [\boldsymbol{V}_1|\ldots|\boldsymbol{V}_m] \in \mathbb{C}^{n\times ms}$, and

$$\mathcal{H}_m = \begin{bmatrix} H_{1,1} & H_{1,2} & \ldots & H_{1,m} \\ H_{2,1} & H_{2,2} & \ldots & H_{2,m} \\ & \ddots & \ddots & \vdots \\ & & H_{m,m-1} & H_{m,m} \end{bmatrix}.$$

Figure 2.1 illustrates the block Arnoldi relation. By construction, each block entry $H_{j,k}$ is an element of $\mathbb{S}$, so it is therefore natural to regard $\mathcal{H}_m$ as an operator which maps $\mathbb{S}^m$, the subspace of $\mathbb{C}^{ms\times s}$ whose block vectors take entries from $\mathbb{S}$, onto itself. We will always do so in this paper and stress this by noting that $\mathcal{H}_m \in \mathbb{S}^{m\times m}$.

How the choice of $\mathbb{S}$ affects the structure of $\mathcal{H}_m$ is depicted in Figure 2.2. In the classical case (top left), $\mathcal{H}_m$ is a block upper Hessenberg matrix, in which the only blocks with a fixed structure are the block lower diagonal elements, which are upper triangular. The global and loop-interchange cases (top right) produce a comparatively sparse pattern. In particular, the global case produces
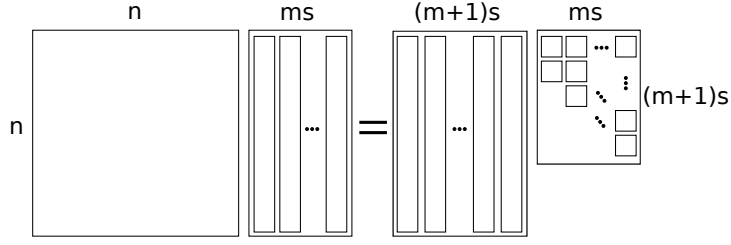
Fig. 2.1: Illustration of the block Arnoldi relation.

$\mathcal{H}_m = H \otimes I_s$, for some $m \times m$ upper Hessenberg matrix H. A hybrid inner product results in a structure whose sparsity is, "in between" that of the classical and loop-interchange structures.

2.1. Matrix polynomials and λ-matrices. Since

$$\mathscr{K}_m^{\mathbb{S}}(A, \boldsymbol{B}) = \text{span}^{\mathbb{S}}\{\boldsymbol{B}, A\boldsymbol{B}, \ldots, A^{m-1}\boldsymbol{B}\} = \left\{ \sum_{k=1}^m A^{k-1}\boldsymbol{B}C_k : C_k \in \mathbb{S} \right\},$$

there is an inherent connection between the Krylov subspace $\mathscr{K}_m^{\mathbb{S}}(A, \boldsymbol{B})$ and polynomials with matrix coefficients.

Let $\{C_j\}_{j=0}^m \subset \mathbb{S}$, with $C_m \neq 0_s$, and let $P : \mathbb{C}^{s \times s} \to \mathbb{C}^{s \times s}$ be given by $P(\Lambda) = \sum_{j=0}^m \Lambda^j C_j$. We denote the space of all such P by $\mathbb{P}_m(\mathbb{S})$ and refer to its elements as *matrix polynomials of degree m*. By $\bar{\mathbb{P}}_m(\mathbb{S})$, we denote the subset of matrix polynomials in $\mathbb{P}_m(\mathbb{S})$ with $C_0 = I_s$. In the special case when $\Lambda = \lambda I_s$ for a scalar λ, $P(\lambda) = \sum_{k=0}^m \lambda^k C_k$ can thus be regarded as a mapping from $\mathbb{C}$ to $\mathbb{C}^{s \times s}$ and is called a *λ-matrix*. For more on the properties of $P(\Lambda)$ and $P(\lambda)$ and how they relate to each other, see [15, 29, 42].

Given a matrix polynomial $P \in \mathbb{P}_m(\mathbb{S})$, one can extract its matrix coefficients $\{C_j\}_{j=0}^m$ and define the action of P on an $n \times n$ matrix A paired with an $n \times s$ block vector $\boldsymbol{B}$ via the operation $\circ$ as follows:

$$P(A) \circ \boldsymbol{B} := \sum_{k=0}^m A^k \boldsymbol{B} C_k. \tag{2.5}$$

In fact, $P(A) \circ \boldsymbol{B}$ could be viewed as a matrix polynomial in its own right with coefficients $\boldsymbol{B}C_k \in \mathbb{C}^{n \times s}$. However, we find it more appropriate to think of the object $P(A) \circ \boldsymbol{B}$ in terms of the underlying matrix polynomial or λ-matrix P with coefficients $C_k \in \mathbb{S}$.

With the operator $\circ$, we can succinctly describe $\mathscr{K}_m^{\mathbb{S}}(A, \boldsymbol{B})$ as

$$\mathscr{K}_m^{\mathbb{S}}(A, \boldsymbol{B}) = \{P(A) \circ \boldsymbol{B} : P \in \mathbb{P}_{m-1}(\mathbb{S})\}. \tag{2.6}$$

3. Block full orthogonalization methods (BFOM). In this section, we recast the convergence theory of block full orthogonalization methods (BFOM) for linear systems in terms of our new generalized framework. To start, recall that $\mathscr{K}_m^{\mathbb{S}}$ is the $\mathbb{S}$-span of the $\langle\!\langle \cdot, \cdot \rangle\!\rangle_{\mathbb{S}}$-orthonormal
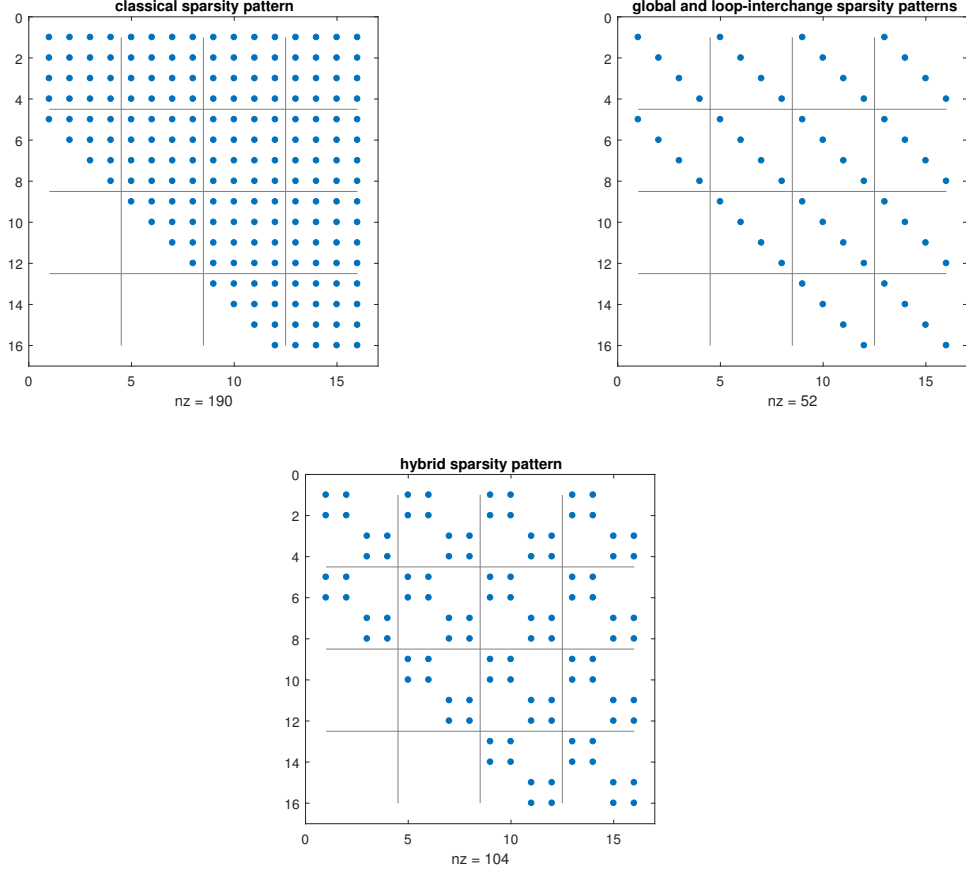
Fig. 2.2: Sparsity patterns of $\mathcal{H}_4$ for different block inner products and $s = 4$, with $q = 2$ for the hybrid example.

Arnoldi basis block vectors $\boldsymbol{V}_1, \ldots, \boldsymbol{V}_m$ which, using $\boldsymbol{\mathcal{V}}_m = [\boldsymbol{V}_1 \mid \cdots \mid \boldsymbol{\mathcal{V}}_m]$ (see (2.4)) can be stated as

$$\mathscr{K}_m^{\mathbb{S}}(A, \boldsymbol{B}) = \{\boldsymbol{\mathcal{V}}_m \boldsymbol{Y} : \boldsymbol{Y} \in \mathbb{S}^m\}. \tag{3.1}$$

In particular, we have $\boldsymbol{B} = \boldsymbol{V}_1 B = \boldsymbol{\mathcal{V}}_m \widehat{\boldsymbol{E}}_1 B$, with the unit block vector $\widehat{\boldsymbol{E}}_1 \in \mathbb{S}^m$ as defined in section 1.4.

Given a block Krylov subspace $\mathscr{K}_m^{\mathbb{S}}(A, \boldsymbol{B})$ and the block inner product $\langle\!\langle \cdot, \cdot \rangle\!\rangle_{\mathbb{S}}$, we now define the mth BFOM approximation to be the block vector $\boldsymbol{X}_m \in \mathscr{K}_m^{\mathbb{S}}(A, \boldsymbol{B})$ satisfying the *block Galerkin condition*

$$\boldsymbol{R}_m := \boldsymbol{B} - A\boldsymbol{X}_m \text{ is } \langle\!\langle \cdot, \cdot \rangle\!\rangle_{\mathbb{S}}\text{-orthogonal to } \mathscr{K}_m^{\mathbb{S}}(A, \boldsymbol{B}). \tag{3.2}$$

THEOREM 3.1. *Assume that $\mathcal{H}_m : \mathbb{S}^m \to \mathbb{S}^m$ is nonsingular and put $\boldsymbol{Y}_m = \mathcal{H}_m^{-1}\widehat{\boldsymbol{E}}_1$. Then*

$$\boldsymbol{X}_m := \boldsymbol{\mathcal{V}}_m \boldsymbol{Y}_m B \tag{3.3}$$

belongs to $\mathscr{K}^{\mathbb{S}}(A, \boldsymbol{B})_m$ and satisfies the block Galerkin condition (3.2).

Proof. Since $B \in \mathbb{S}$ and $\boldsymbol{Y}_m \in \mathbb{S}^m$, we have that $\boldsymbol{Y}_m B \in \mathbb{S}^m$ and thus $\boldsymbol{X}_m \in \mathscr{K}^{\mathbb{S}}(A, \boldsymbol{B})_m$ by (3.1). Using the block Arnoldi relation (2.4) and the fact that $\boldsymbol{B} = \boldsymbol{\mathcal{V}}_m \widehat{\boldsymbol{E}}_1 B$, we have

$$
\begin{aligned}
\boldsymbol{R}_m \;=\; \boldsymbol{B} - A\boldsymbol{X}_m &= \boldsymbol{\mathcal{V}}_m \widehat{\boldsymbol{E}}_1 B - A\boldsymbol{\mathcal{V}}_m \boldsymbol{Y}_m B \\
&= \boldsymbol{\mathcal{V}}_m \widehat{\boldsymbol{E}}_1 B - (\boldsymbol{\mathcal{V}}_m \mathcal{H}_m + \boldsymbol{V}_{m+1} H_{m+1,m} \widehat{\boldsymbol{E}}_m^*) \boldsymbol{Y}_m B \\
&= -\boldsymbol{V}_{m+1} H_{m+1,m} \widehat{\boldsymbol{E}}_m^* \boldsymbol{Y}_m B \\
&= \boldsymbol{V}_{m+1} C_m \quad \text{with } C_m := -H_{m+1,m} \widehat{\boldsymbol{E}}_m^* \boldsymbol{Y}_m B \in \mathbb{S}.
\end{aligned}
\tag{3.4}
$$

Thus, since $\boldsymbol{V}_{m+1}$ is $\langle\!\langle \cdot, \cdot \rangle\!\rangle_{\mathbb{S}}$-orthogonal to $\mathscr{K}_m^{\mathbb{S}}(A, \boldsymbol{B})$ by construction, so is $\boldsymbol{V}_{m+1} C_m$, implying that (3.3) satisfies (3.2). $\square$

We say that $\boldsymbol{R}_m$ and $\boldsymbol{V}_{m+1}$ are *cospatial*, since, by (3.4) the columns of each span the same s-dimensional subspace in $\mathbb{C}^n$.

It is also helpful to consider the polynomial representations of $\boldsymbol{X}_m$ and $\boldsymbol{R}_m$. Since $\boldsymbol{X}_m \in \mathscr{K}_m^{\mathbb{S}}(A, \boldsymbol{B})$, we know there exists Q_{m-1} in $\mathbb{P}_{m-1}(\mathbb{S})$ such that $\boldsymbol{X}_m = Q_{m-1}(A) \circ \boldsymbol{B}$. Then

$$
\begin{aligned}
\boldsymbol{R}_m = \boldsymbol{B} - A\boldsymbol{X}_m &= \boldsymbol{B} - A(Q_{m-1}(A) \circ \boldsymbol{B}) = A^0 \boldsymbol{B} I_s - (AQ_{m-1}(A)) \circ \boldsymbol{B} \\
&= P_m(A) \circ \boldsymbol{B},
\end{aligned}
\tag{3.5}
$$

where $P_m(\lambda) := I_s - \lambda Q_{m-1}(\lambda) \in \bar{\mathbb{P}}_m(\mathbb{S})$.

3.1. Convergence results for BFOM. We analyze the convergence of BFOM with a general block inner product $\langle\!\langle \cdot, \cdot \rangle\!\rangle_{\mathbb{S}}$ in the case that the matrix A is $\langle\!\langle \cdot, \cdot \rangle\!\rangle_{\mathbb{S}}$-self-adjoint and positive definite in a sense to be defined shortly. Our results reduce to what is well-known for BCG in the case of the classical block inner product [49]. For the global and loop-interchange block inner products, the results are directly related to what is known for CG for a single right-hand side, and we briefly discuss this at the end of the section.

Assume that A is HPD and let $0 < \lambda_{\min} \le \lambda_{\max}$ denote the smallest and largest eigenvalues of A, respectively. The following constants play an important role:

$$\kappa := \frac{\lambda_{\max}}{\lambda_{\min}}, \quad c := \frac{\sqrt{\kappa} - 1}{\sqrt{\kappa} + 1}, \quad \text{and} \quad \xi_m := \frac{1}{\cosh(m \ln c)} = \frac{2}{c^m + c^{-m}}. \tag{3.6}$$

If $\kappa = 1$, then we set $\xi_m = 0$. Our goal is to mimic the well-known convergence result for (non-block) CG stated in the following theorem; see, e.g., [44, Ch. 8] and [54, Ch. 6]. This result uses the A-norm $\|\boldsymbol{x}\|_A := \langle \boldsymbol{x}, A\boldsymbol{x} \rangle_2^{1/2}$.

THEOREM 3.2. *The error $\boldsymbol{e}_m$ for the CG iterate at step m satisfies*

$$\|\boldsymbol{e}_m\|_A = \min_{\boldsymbol{x} \in \mathscr{K}_m(A, \boldsymbol{b})} \|\boldsymbol{x}_* - \boldsymbol{x}\|_A \le \xi_m \|\boldsymbol{e}_0\|_A \le 2c^m \|\boldsymbol{e}_0\|_A.$$

3.1.1. A scalar inner product and norm. To deduce results like those of Theorem 3.2 in our block framework, we introduce a scalar inner product $\langle \cdot, \cdot \rangle_{\mathbb{S}} : \mathbb{C}^{n \times s} \times \mathbb{C}^{n \times s} \to \mathbb{C}$:

$$\langle \boldsymbol{X}, \boldsymbol{Y} \rangle_{\mathbb{S}} := \operatorname{trace} \langle\!\langle \boldsymbol{Y}, \boldsymbol{X} \rangle\!\rangle_{\mathbb{S}}. \tag{3.7}$$

The properties of $\langle\!\langle \cdot, \cdot \rangle\!\rangle_{\mathbb{S}}$ from Definition 2.1 ensure that (3.7) is a true inner product on $\mathbb{C}^{n \times s}$, which further induces the norm

$$\|\boldsymbol{X}\|_{\mathbb{S}} := \langle \boldsymbol{X}, \boldsymbol{X} \rangle_{\mathbb{S}}^{\frac{1}{2}}.$$

We note that in the case of Examples 2.3 and 2.5, the resulting norms $\|\cdot\|_{\mathbb{S}}^{\mathrm{Cl}}$ and $\|\cdot\|_{\mathbb{S}}^{\mathrm{Li}}$ are both identical to the Fobenius norm $\|\cdot\|_{\mathrm{F}}$,

while in the case of Example 2.4, we have that $\|\cdot\|_{\mathbb{S}}^{\mathrm{Gl}} = s \, \|\cdot\|_{\mathrm{F}}$.

DEFINITION 3.3. *A matrix $A \in \mathbb{C}^{n \times n}$, as an operator on $\mathbb{C}^{n \times s}$, is called*

(i) $\langle \cdot, \cdot \rangle_{\mathbb{S}}$-self-adjoint if for all $\boldsymbol{X}, \boldsymbol{Y} \in \mathbb{C}^{n \times s}$ we have $\langle A\boldsymbol{X}, \boldsymbol{Y} \rangle_{\mathbb{S}} = \langle \boldsymbol{X}, A\boldsymbol{Y} \rangle_{\mathbb{S}}$,

(ii) $\langle \cdot, \cdot \rangle_{\mathbb{S}}$-positive definite if, in addition, for all nonzero $\boldsymbol{X} \in \mathbb{C}^{n \times s}$ we have $\langle \boldsymbol{X}, A\boldsymbol{X} \rangle_{\mathbb{S}} > 0$.

REMARK 3.4. The following assertions follow immediatley from the definition.

(i) If A is $\langle\!\langle \cdot, \cdot \rangle\!\rangle_{\mathbb{S}}$-self-adjoint then A is also $\langle \cdot, \cdot \rangle_{\mathbb{S}}$-self-adjoint. The converse does not necessarily hold.

(ii) For A $\langle \cdot, \cdot \rangle_{\mathbb{S}}$-positive-definite, $\operatorname{spec}(A) \subset (0, \infty)$.

As for (ii), we note that whether we regard $A \in \mathbb{C}^{n \times n}$ as an operator on $\mathbb{C}^n$ or on $\mathbb{C}^{n \times s}$, it has the same spectrum, albeit the geometric multiplicities of the eigenvalues are multiplied by s for the latter.

LEMMA 3.5. *Let p be a scalar-valued polynomial with real coefficients and let A be $\langle \cdot, \cdot \rangle_{\mathbb{S}}$-self-adjoint. Then*

$$\|p(A)\|_{\mathbb{S}} = \max_{\lambda \in \operatorname{spec}(A)} |p(\lambda)| .$$

Proof. As a $\langle \cdot, \cdot \rangle_{\mathbb{S}}$-self-adjoint operator on $\mathbb{C}^{n \times s}$, A admits an $\langle \cdot, \cdot \rangle_{\mathbb{S}}$-orthonormal basis of eigenvectors from $\mathbb{C}^{n \times s}$, implying that its operator norm $\|A\|_{\mathbb{S}}$ is given as

$$\|A\|_{\mathbb{S}} = \max_{\lambda \in \operatorname{spec}(A)} |\lambda| .$$

Since p has real coefficients, $p(A)$ is also $\langle \cdot, \cdot \rangle_{\mathbb{S}}$-self-adjoint, and $\operatorname{spec}(p(A)) = \{p(\lambda) : \lambda \in \operatorname{spec}(A)\}$. This directly gives $\|p(A)\|_{\mathbb{S}} = \max_{\lambda \in \operatorname{spec}(A)} |p(\lambda)|$. $\square$

When A is $\langle \cdot, \cdot \rangle_{\mathbb{S}}$-positive definite, we have for any rational function g, which is defined and positive on $(0, \infty)$, a $g(A)$-weighted inner product and norm, defined as

$$\langle \boldsymbol{X}, \boldsymbol{Y} \rangle_{\mathbb{S}\text{-}g(A)} := \langle \boldsymbol{X}, g(A)\boldsymbol{Y} \rangle_{\mathbb{S}}, \quad \|\boldsymbol{X}\|_{\mathbb{S}\text{-}g(A)} := \langle \boldsymbol{X}, \boldsymbol{X} \rangle_{\mathbb{S}\text{-}g(A)}^{\frac{1}{2}}.$$

We will particularly need the cases $g(A) = A^{-1}$ and $g(A) = (A + tI)$, $t \geq 0$, in Theorem 4.5. Lemma 3.5 carries over to the $g(A)$-weighted inner products.

LEMMA 3.6. *Let p be a scalar-valued polynomial with real coefficients, $g : (0, \infty) \to (0, \infty)$ a rational function, and A be $\langle \cdot, \cdot \rangle_{\mathbb{S}}$-positive-definite. Then*

$$\|p(A)\boldsymbol{X}\|_{\mathbb{S}\text{-}g(A)} = \max_{\lambda \in \operatorname{spec}(A)} |p(\lambda)|. \tag{3.8}$$

Proof. Since A is $\langle \cdot, \cdot \rangle_{\mathbb{S}}$-self-adjoint, it also is $\langle \cdot, \cdot \rangle_{\mathbb{S}\text{-}g(A)}$-self-adjoint due to

$$\langle A\boldsymbol{X}, \boldsymbol{Y} \rangle_{\mathbb{S}\text{-}g(A)} = \langle A\boldsymbol{X}, g(A)\boldsymbol{Y} \rangle_{\mathbb{S}} = \langle \boldsymbol{X}, Ag(A)\boldsymbol{Y} \rangle_{\mathbb{S}} = \langle \boldsymbol{X}, g(A)A\boldsymbol{Y} \rangle_{\mathbb{S}} = \langle \boldsymbol{X}, A\boldsymbol{Y} \rangle_{\mathbb{S}\text{-}g(A)}.$$

Application of Lemma 3.5 for the $\langle \cdot, \cdot \rangle_{\mathbb{S}\text{-}g(A)}$ inner product leads to the desired result. □

The scalar inner product $\langle \cdot, \cdot \rangle_{\mathbb{S}}$ induces a traditional, scalar notion of $\langle \cdot, \cdot \rangle_{\mathbb{S}}$-orthogonality, and similarly for the $g(A)$-weighted versions. Trivially, $\langle\!\langle \cdot, \cdot \rangle\!\rangle_{\mathbb{S}}$-orthogonality implies $\langle \cdot, \cdot \rangle_{\mathbb{S}}$-orthogonality, since $\langle\!\langle \boldsymbol{Y}, \boldsymbol{X} \rangle\!\rangle_{\mathbb{S}} = 0_s$ implies $\langle \boldsymbol{X}, \boldsymbol{Y} \rangle_{\mathbb{S}} = \text{trace}\left(\langle\!\langle \boldsymbol{Y}, \boldsymbol{X} \rangle\!\rangle_{\mathbb{S}}\right) = 0$, and likewise for the $g(A)$-weighted cases. The converse, however, is not true. For example, consider the classical block inner product (Example 2.3) on $\mathbb{C}^{3 \times 2}$, let $\boldsymbol{v} := [1, 0, 0]^T$, $\boldsymbol{w} := [0, 1, 0]^T$ and define $\boldsymbol{Y} := [\boldsymbol{v}|\boldsymbol{w}]$, $\boldsymbol{X} := [\boldsymbol{w}|\boldsymbol{v}]$. Then

$$\langle\!\langle \boldsymbol{Y}, \boldsymbol{X} \rangle\!\rangle_{\mathbb{S}}^{\mathrm{Cl}} = \begin{pmatrix} \boldsymbol{v}^*\boldsymbol{w} & \boldsymbol{v}^*\boldsymbol{v} \\ \boldsymbol{w}^*\boldsymbol{w} & \boldsymbol{w}^*\boldsymbol{v} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

implying that $\boldsymbol{X}$ and $\boldsymbol{Y}$ are not block orthogonal, while $\text{trace} \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right) = 0$, implying that $\boldsymbol{X}$ and $\boldsymbol{Y}$ are $\langle \cdot, \cdot \rangle_{\mathbb{S}}^{\mathrm{Cl}}$-orthogonal.

3.1.2. A variational characterization and error bounds. We denote by $\boldsymbol{\mathcal{P}}$ the $\langle \cdot, \cdot \rangle_{\mathbb{S}\text{-}A}$-orthogonal projector onto a subspace $\mathscr{K}$ of $\mathbb{C}^{n \times s}$. It then follows that for any $\boldsymbol{Y} \in \mathbb{C}^{n \times s}$,

$$\|\boldsymbol{Y} - \boldsymbol{\mathcal{P}}\boldsymbol{Y}\|_{\mathbb{S}\text{-}A} = \min_{\boldsymbol{X} \in \mathscr{K}} \|\boldsymbol{Y} - \boldsymbol{X}\|_{\mathbb{S}\text{-}A},$$

and that

$$\|\boldsymbol{Y} - \boldsymbol{Z}\|_{\mathbb{S}\text{-}A} = \min_{\boldsymbol{X} \in \mathscr{K}} \|\boldsymbol{Y} - \boldsymbol{X}\|_{\mathbb{S}\text{-}A} \iff \boldsymbol{Z} \in \mathscr{K} \text{ and } \boldsymbol{Y} - \boldsymbol{Z} \text{ is } \langle \cdot, \cdot \rangle_{\mathbb{S}\text{-}A}\text{-orthogonal to } \mathscr{K}. \quad (3.9)$$

We are now in a position to formulate the following generalization of Theorem 3.2.

THEOREM 3.7. *Let $A \in \mathbb{C}^{n \times n}$ be $\langle\!\langle \cdot, \cdot \rangle\!\rangle_{\mathbb{S}}$-self-adjoint and $\langle \cdot, \cdot \rangle_{\mathbb{S}}$-positive definite, and $\boldsymbol{B} \in \mathbb{C}^{n \times s}$ be a block right-hand-side vector. Then the BFOM error $\boldsymbol{E}_m = \boldsymbol{X}_* - \boldsymbol{X}_m$ satisfies*

$$\|\boldsymbol{E}_m\|_{\mathbb{S}\text{-}A} = \min_{\boldsymbol{X} \in \mathscr{K}_m^{\mathbb{S}}(A, \boldsymbol{B})} \|\boldsymbol{X}_* - \boldsymbol{X}\|_{\mathbb{S}\text{-}A} \leq \xi_m \|\boldsymbol{E}_0\|_{\mathbb{S}\text{-}A}, \quad (3.10)$$

with ξ_m from (3.6).

Proof. Since $\boldsymbol{X}_m$ satisfies the block Galerkin condition (3.2), $\boldsymbol{R}_m$ is $\langle\!\langle \cdot, \cdot \rangle\!\rangle_{\mathbb{S}}$-orthogonal and, consequently, $\langle \cdot, \cdot \rangle_{\mathbb{S}}$-orthogonal to $\mathscr{K}_m^{\mathbb{S}}(A, \boldsymbol{B})$. Then for all $\boldsymbol{V} \in \mathscr{K}_m^{\mathbb{S}}(A, \boldsymbol{B})$,

$$0 = \langle \boldsymbol{R}_m, \boldsymbol{V} \rangle_{\mathbb{S}} = \langle A\boldsymbol{E}_m, \boldsymbol{V} \rangle_{\mathbb{S}} = \langle \boldsymbol{E}_m, \boldsymbol{V} \rangle_{\mathbb{S}\text{-}A}.$$

Since $\boldsymbol{E}_m = \boldsymbol{X}_* - \boldsymbol{X}_m$ with $\boldsymbol{X}_m \in \mathscr{K}_m^{\mathbb{S}}(A, \boldsymbol{B})$, applying (3.9) gives the equality in (3.10). To prove the inequality in (3.10), we make use of the polynomial characterization of the block Krylov subspace $\mathscr{K}_m^{\mathbb{S}}(A, \boldsymbol{B})$ given in (2.6). We thus have that

$$\|P_m(A) \circ \boldsymbol{E}_0\|_{\mathbb{S}\text{-}A} = \min_{P \in \bar{\mathbb{P}}_m(\mathbb{S})} \|P(A) \circ \boldsymbol{E}_0\|_{\mathbb{S}\text{-}A}, \quad (3.11)$$

where P_m is the BFOM residual polynomial as in (3.5). By the embedding $p(\lambda) = 1 + \sum_{i=1}^m \gamma_i \lambda^i \hookrightarrow P_p(\lambda) = I_s + \sum_{i=1}^m (\gamma_i I_s)\lambda^i$, we can regard $\bar{\mathbb{P}}_m(\mathbb{C})$ as a subspace of $\bar{\mathbb{P}}_m(\mathbb{S})$, with $P_p(A) \circ \boldsymbol{X} = p(A)\boldsymbol{X}$. Along with (3.11) and Lemma 3.6 this gives that

$$\|P_m(A) \circ \boldsymbol{E}_0\|_{\mathbb{S}\text{-}A} \leq \|p(A)\boldsymbol{E}_0\|_{\mathbb{S}\text{-}A} \leq \max_{\lambda \in \text{spec}(A)} |p(\lambda)| \cdot \|\boldsymbol{E}_0\|_{\mathbb{S}\text{-}A} \text{ for any } p \in \bar{\mathbb{P}}_m(\mathbb{C}).$$

If we now take p as the (scaled) Chebyshev polynomial of degree m for the interval $[\lambda_{\min}, \lambda_{\max}]$ (as in, e.g., [54, Chapter 6]), then $\max_{\lambda \in [\lambda_{\min}, \lambda_{\max}]} |p(\lambda)| \leq \xi_m$, leading to the inequality in (3.10). $\square$

To put Theorem 3.7 in perspective, consider the specific block inner products $\langle\!\langle \cdot, \cdot \rangle\!\rangle_{\mathbb{S}}$ from Examples 2.3-2.5. In the classical and loop interchange cases, $\|\boldsymbol{X}\|_{\mathbb{S}\text{-}A} = \|\boldsymbol{X}\|_{A\text{-F}} = \sqrt{\text{trace}(\boldsymbol{X}^* A \boldsymbol{X})}$, and in the global case, $\|\boldsymbol{X}\|_{\mathbb{S}\text{-}A} = s \|\boldsymbol{X}\|_{A\text{-F}}$. Thus, in all three cases, Theorem 3.7 gives

$$\|\boldsymbol{E}_m\|_{A\text{-F}} \leq \xi_m \|\boldsymbol{E}_0\|_{A\text{-F}}. \tag{3.12}$$

For the classical case, this result is contained in unpublished work by Eisenstat [20], who rewrites results from [49] in terms of the A-weighted Frobenius norm. In the loop interchange case, we can use the standard CG error bound from Theorem 3.2 for each column as an alternative way to arrive at (3.12). In the global case, the estimate (3.12) can also be obtained as follows. Let $\mathcal{A} = I_s \otimes A$. Then solving the block linear system $A\boldsymbol{X} = \boldsymbol{B}$ with GlBFOM is identical to solving $\mathcal{A}\text{vec}(\boldsymbol{X}) = \text{vec}(\boldsymbol{B})$ with FOM [36, Theorem 1], where vec is the operator that reshapes an $n \times s$ block vector into an $ns \times 1$ vector. Since $\mathcal{A}$ and A have identical spectra, κ, c, and ξ_m are just as in (3.6). Applying Theorem 3.2 we obtain that

$$\|\text{vec}(\boldsymbol{E}_m)\|_{\mathcal{A}} \leq \xi_m \|\text{vec}(\boldsymbol{E}_0)\|_{\mathcal{A}}.$$

Converting everything back to block form gives (3.12).

It is worth mentioning that since the *-algebras for the three examples are nested as $\mathbb{S}^{\text{Gl}} \subseteq \mathbb{S}^{\text{Li}} \subseteq \mathbb{S}^{\text{Cl}}$, the variational characterization from Theorem 3.7 directly provides the comparison

$$\left\|\boldsymbol{E}_m^{\text{Cl}}\right\|_{A\text{-F}} \leq \left\|\boldsymbol{E}_m^{\text{Li}}\right\|_{A\text{-F}} \leq \left\|\boldsymbol{E}_m^{\text{Gl}}\right\|_{A\text{-F}}. \tag{3.13}$$

4. BFOM for functions of matrices: B(FOM)^2. We begin this section by considering a generic function $f : \mathbb{D} \subset \mathbb{C} \to \mathbb{C}$ and an $n \times n$ matrix A, along with a starting block vector $\boldsymbol{B}$, a block inner product $\langle\!\langle \cdot, \cdot \rangle\!\rangle_{\mathbb{S}}$ on $\mathbb{C}^{n \times s}$ with scaling quotient N, and corresponding outputs from the Arnoldi process (Algorithm 1) fulfilling the block Arnoldi relation (2.4). Assuming $f(A)\boldsymbol{B}$ exists, we define the B(FOM)^2 approximation to $f(A)\boldsymbol{B}$ as

$$\boldsymbol{F}_m := \boldsymbol{\mathcal{V}}_m f(\mathcal{H}_m) \widehat{\boldsymbol{E}}_1 B, \tag{4.1}$$

where, just as $\mathcal{H}_m$, the matrix function $f(\mathcal{H}_m)$ is considered as an operator from $\mathbb{S}^m$ to $\mathbb{S}^m$.

If we can express f as a contour integral of the form

$$f(z) = \int_\Gamma \frac{g(t)}{z + t}\, dt, \tag{4.2}$$

then[2] $f(A)\boldsymbol{B} = \int_\Gamma g(t)(A + tI)^{-1}\boldsymbol{B}$ and

$$\boldsymbol{F}_m = \int_\Gamma g(t) \boldsymbol{\mathcal{V}}_m (\mathcal{H}_m + tI)^{-1} \widehat{\boldsymbol{E}}_1 B. \tag{4.3}$$

A crucial connection to BFOM for systems of equations now arises. The block Arnoldi algorithm is *shift invariant*, i.e., the block Arnoldi procedure for $A + tI$ produces the same basis $\boldsymbol{\mathcal{V}}_m$ as the

---

[2]We refer to, e.g., [27, 37] for a discussion of why and when this definition of a matrix function coincides with the standard definition based on interpolating polynomials.

process for A, as well as the same block Arnoldi relation, but with the matrix $\mathcal{H}_m$ replaced by $\mathcal{H}_m + tI$. Therefore, for each $t \in \Gamma$, the block vector $\boldsymbol{\mathcal{V}}_m(\mathcal{H}_m + tI)^{-1}\widehat{\boldsymbol{E}}_1 B$ appearing in (4.3) is precisely the mth BFOM approximation to the solution $\boldsymbol{X}(t)$ of the block linear system $(A + tI)\boldsymbol{X}(t) = \boldsymbol{B}$.

Note that $f(\mathcal{H}_m)$ is not necessarily defined when $f(A)$ is, because $\mathcal{H}_m$ could have an eigenvalue on which f is not defined or not sufficiently smooth. We now show that if f is defined and sufficiently smooth on an appropriate superset of $\mathrm{spec}(A)$, then $f(\mathcal{H}_m)$ is defined as well. Recall that for an operator H on a Hilbert space T with inner product $\langle \cdot, \cdot \rangle$, the field of values $\mathbb{F}_{\langle \cdot, \cdot \rangle}(H)$ is given as

$$\mathbb{F}_{\langle \cdot, \cdot \rangle}(H) = \left\{ \frac{\langle Hx, x \rangle}{\langle x, x \rangle} : x \in T, x \neq 0 \right\},$$

and that the field of values contains the spectrum of H. We use the field of values for A as an operator on $\mathbb{C}^{n \times s}$ with the inner product $\langle \cdot, \cdot \rangle_{\mathbb{S}}$, and for $\mathcal{H}_m$ as an operator on $\mathbb{S}^m$ with the inner product $\langle \cdot, \cdot \rangle_{\boldsymbol{\mathcal{V}}_m}$ defined as

$$\langle \boldsymbol{X}, \boldsymbol{Y} \rangle_{\boldsymbol{\mathcal{V}}_m} := \langle \boldsymbol{\mathcal{V}}_m \boldsymbol{X}, \boldsymbol{\mathcal{V}}_m \boldsymbol{Y} \rangle_{\mathbb{S}}, \quad \boldsymbol{X}, \boldsymbol{Y} \in \mathbb{S}^m,$$

where $\boldsymbol{\mathcal{V}}_m$ is the matrix of the block Arnoldi vectors; see (2.4). The following lemma relates the two field of values.

LEMMA 4.1. *We have*

$$\mathbb{F}_{\langle \cdot, \cdot \rangle_{\mathbb{S}}}(\mathcal{H}_m) \subset \mathbb{F}_{\langle \cdot, \cdot \rangle_{\mathbb{S}}}(A).$$

Proof. We first observe that by the Arnoldi relation (2.4) we have that

$$\begin{aligned}
\langle \boldsymbol{X}, \mathcal{H}_m \boldsymbol{X} \rangle_{\boldsymbol{\mathcal{V}}_m} &= \langle \boldsymbol{\mathcal{V}}_m \boldsymbol{X}, \boldsymbol{\mathcal{V}}_m \mathcal{H}_m \boldsymbol{X} \rangle_{\mathbb{S}} \\
&= \langle \boldsymbol{\mathcal{V}}_m \boldsymbol{X}, A\boldsymbol{\mathcal{V}}_m \boldsymbol{X} \rangle_{\mathbb{S}} - \langle \boldsymbol{\mathcal{V}}_m \boldsymbol{X}, \boldsymbol{V}_{m+1} H_{m+1,m} \widehat{\boldsymbol{E}}_m^* \boldsymbol{X} \rangle_{\mathbb{S}} \\
&= \langle \boldsymbol{\mathcal{V}}_m \boldsymbol{X}, A\boldsymbol{\mathcal{V}}_m \boldsymbol{X} \rangle_{\mathbb{S}},
\end{aligned}$$

where the last equality holds since $\langle\!\langle \boldsymbol{\mathcal{V}}_m \boldsymbol{X}, \boldsymbol{V}_{m+1} H_{m+1,m} \widehat{\boldsymbol{E}}_m^* \boldsymbol{X} \rangle\!\rangle_{\mathbb{S}} = 0_s$, which one can see by breaking $\boldsymbol{\mathcal{V}}_m \boldsymbol{X}$ into components and applying the fact that the $\boldsymbol{V}_j$ are orthonormal. Clearly, $\boldsymbol{\mathcal{V}}_m \boldsymbol{X} \in \mathbb{C}^{n \times s}$. Since, moreover, $\langle \boldsymbol{\mathcal{V}}_m \boldsymbol{X}, \boldsymbol{\mathcal{V}}_m \boldsymbol{X} \rangle_{\mathbb{S}} = \langle \boldsymbol{X}, \boldsymbol{X} \rangle_{\boldsymbol{\mathcal{V}}_m}$, this gives the desired result. $\square$

As a direct consequence, we obtain the following sufficient condition for the existence of $f(\mathcal{H}_m)$.

THEOREM 4.2. *If f is defined and sufficiently smooth on $\mathbb{F}_{\langle \cdot, \cdot \rangle_{\mathbb{S}}}(A)$, then $f(\mathcal{H}_m) : \mathbb{S}^m \to \mathbb{S}^m$ is well defined, and consequently, so is the approximation $\boldsymbol{F}_m$ given by (4.1).*

In the remainder of this section we concentrate on Cauchy-Stieltjes functions as a special class of functions closely related to those of the form (4.2) and develop a restart approach for B(FOM)^2 on such functions. It is important to note that, in principle, the B(FOM)^2 approximation (4.1) and restart technique can be used for any f of the form (4.2), particularly analytic functions. We, however, develop theory for Stieltjes functions and $\langle\!\langle \cdot, \cdot \rangle\!\rangle_{\mathbb{S}}$-self-adjoint, $\langle \cdot, \cdot \rangle_{\mathbb{S}}$-positive definite matrices A only.

4.1. Stieltjes functions. A Stieltjes or Cauchy-Stieltjes function is a function $f : \mathbb{C} \backslash (-\infty, 0] \to \mathbb{C}$ that can be written as a Riemann-Stieltjes integral as follows:

$$f(z) = \int_0^\infty \frac{1}{z + t} \, \mathrm{d}\mu(t), \tag{4.4}$$

where μ is monotonically increasing and nonnegative on $[0, \infty)$ with $\int_0^\infty \frac{1}{t+1} \, d\mu(t) < \infty$. The relation between (4.4) and (4.2) becomes obvious if μ is differentiable, since then $g = \mu'$. Stieltjes functions are a particular case of Markov-type functions [24, 41]. The matrix Stieltjes function $f(A)$ is defined if A has no eigenvalues on $(-\infty, 0]$. If $\mathbb{F}_{\langle \cdot, \cdot \rangle_{\mathbb{S}}}(A) \cap (-\infty, 0] = \emptyset$, then by Theorem 4.2 the B(FOM)^2 approximation $\boldsymbol{F}_m = \boldsymbol{\mathcal{V}}_m f(\mathcal{H}_m) \widehat{\boldsymbol{E}}_1 B$ exists.

Many important functions are Stieltjes functions. For example, for $\alpha \in (0, 1)$,

$$z^{-\alpha} = \frac{\sin((1 - \alpha)\pi)}{\pi} \int_0^\infty \frac{1}{z + t} \, d\mu(t), \qquad \text{with} \quad d\mu(t) = t^{-\alpha} \, dt, \tag{4.5}$$

and

$$\frac{\log(1 + z)}{z} = \int_0^\infty \frac{1}{z + t} \, d\mu(t), \quad \text{with} \quad d\mu(t) = \begin{cases} 0 \, dt & 0 \leq t \leq 1 \\ t^{-1} \, dt & t > 1 \end{cases}. \tag{4.6}$$

For more information on Stieltjes functions, see , e.g., [35].

4.2. Restarts. The computation of a sufficiently accurate $\boldsymbol{F}_m$ faces many limitations in practice, since Krylov methods tend to suffer from steep storage requirements as the number of basis vectors m increases, and even more so in the case of block bases. For linear systems, restarts are the standard approach for mitigating this issue. Based on work for the non-block case in [25], we develop a numerically stable and efficient restart approach for B(FOM)^2. The key is to use the integral representation for f to develop an integral representation for the error of the B(FOM)^2 approximation $\boldsymbol{F}_m$. As we will see, this integral representation is defined in a fundamentally different way compared to the non-block case, but we can nevertheless apply a variant of B(FOM)^2 to it to approximate the error. The error approximation is ultimately defined in a recursive fashion and updated each time a new block basis is computed. Note that the norm of the error approximation can be used as a stopping criterion, an important issue when approximating matrix functions.

Updating the function approximation with an error approximation is the only established approach in the literature for restarting (FOM)^2 for a single vector [1, 19, 25, 39]. Modifications to this technique include varying cycle lengths or choosing more than one block vector for generating the next basis, as is done with thick restarts [24] or recycling [50]. However, for the sake of simplicity, we assume that the number of basis vectors computed per cycle m is constant, and that only one block vector, specifically the last basis vector $\boldsymbol{V}_{m+1}^{(k)}$ from the previous cycle, is used as the starting vector for computing the basis vectors of the next cycle.

4.2.1. Restarted BFOM for shifted linear systems. Essential to the development of a restart procedure for B(FOM)^2 is theory for restarted BFOM for the shifted linear systems $(A + tI)\boldsymbol{X}_*(t) = \boldsymbol{B}$, $t \geq 0$, which we now expound. We begin by fixing m, the number of block basis vectors that are generated and stored at a time in the block Arnoldi process, and henceforth refer to it as the *cycle length*. Using a superscript (k) to index quantities from a given cycle, and prescribing $\boldsymbol{X}_m^{(0)}(t) := \boldsymbol{0}$, the restarted BFOM approximation to $\boldsymbol{X}_*(t)$ obtained after the $k + 1$st cycle is given as

$$\boldsymbol{X}_m^{(k+1)}(t) := \boldsymbol{X}_m^{(k)}(t) + \boldsymbol{Z}_m^{(k)}(t), \quad k = 0, 1, \dots \tag{4.7}$$

with $\boldsymbol{Z}_m^{(k)}(t)$ defined as the BFOM approximation to $\boldsymbol{Z}_*^{(k)}(t)$ in the block residual equation

$$(A + tI)\boldsymbol{Z}_*^{(k)}(t) = \boldsymbol{R}_m^{(k)}(t), \text{ with } \boldsymbol{R}_m^{(k)}(t) := \boldsymbol{B} - (A + tI)\boldsymbol{X}_m^{(k)}(t).$$

Consider the first cycle. By the shift invariance of the block Arnoldi relation (2.4) and by the cospatial relation (3.4), we have that

$$\boldsymbol{R}_m^{(1)}(t) = \boldsymbol{V}_{m+1}^{(1)} C_m^{(1)}(t), \quad \text{with}$$

$$C_m^{(1)}(t) := -H_{m+1,m}^{(1)} \widehat{\boldsymbol{E}}_m^* \boldsymbol{Y}_m^{(1)}(t) B \quad \text{and} \quad \boldsymbol{Y}_m^{(1)}(t) := (\mathcal{H}_m + tI)^{-1} \widehat{\boldsymbol{E}}_1,$$

where the cospatiality factor $C_m^{(1)}(t)$ depends on t, while the block basis vector $\boldsymbol{V}_{m+1}^{(1)}$ does not.

Instead of starting the second cycle with $\boldsymbol{R}_m^{(1)}(t)$, we can just as well start it with $\boldsymbol{V}_{m+1}^{(1)}$, since the two are cospatial. We then obtain the block basis $\{\boldsymbol{V}_1^{(2)} = \boldsymbol{V}_{m+1}^{(1)}, \ldots, \boldsymbol{V}_{m+1}^{(2)}\}$, which block spans $\mathcal{K}_m^{\mathbb{S}}(A, \boldsymbol{V}_1^{(2)})$ and are again independent of t. If we then take

$$\boldsymbol{Z}_m^{(1)}(t) := \boldsymbol{\mathcal{V}}_m^{(2)} \boldsymbol{Y}_m^{(2)}(t) C_m^{(1)}(t) \text{ with } \boldsymbol{Y}_m^{(2)}(t) := \left(\mathcal{H}_m^{(2)} + tI\right)^{-1} \widehat{\boldsymbol{E}}_1,$$

we see, again by the cospatial relation (3.4), that the residual to the equation $(A+tI)\boldsymbol{Z}_*(t) = \boldsymbol{R}_m^{(1)}(t)$ for the approximation $\boldsymbol{Z}^{(1)}(t)$ satisfies

$$\boldsymbol{R}_m^{(1)}(t) - (A + tI)\boldsymbol{Z}_m^{(1)}(t) = -\boldsymbol{V}_{m+1}^{(2)} H_{m+1,m}^{(2)} \widehat{\boldsymbol{E}}_m^* \boldsymbol{Y}_m^{(2)}(t) C_m^{(1)}(t),$$

showing that the residual $\boldsymbol{R}_m^{(1)}(t) - (A + tI)\boldsymbol{Z}_m^{(1)}(t)$ is $\lang\!\langle \cdot, \cdot \rangle\!\rangle_{\mathbb{S}}$-orthogonal to $\mathcal{K}_m^{\mathbb{S}}(A, \boldsymbol{V}_1^{(2)})$, thus satisfying the block Galerkin condition (3.2). This implies that $\boldsymbol{Z}_m^{(1)}(t)$ is indeed the BFOM approximation for the residual equation $(A + tI)\boldsymbol{Z}_*^{(1)}(t) = \boldsymbol{R}_m^{(1)}(t)$. The residual $\boldsymbol{R}_m^{(2)}(t)$ of the updated approximation $\boldsymbol{X}_m^{(2)}(t) = \boldsymbol{X}_m^{(1)}(t) + \boldsymbol{Z}_m^{(1)}(t)$ is then given as

$$\boldsymbol{R}_m^{(2)}(t) = \boldsymbol{R}_m^{(1)}(t) - (A + tI)\boldsymbol{Z}_m^{(1)}(t) = -\boldsymbol{V}_{m+1}^{(2)} H_{m+1,m}^{(2)} \widehat{\boldsymbol{E}}_m^* \boldsymbol{Y}_m^{(2)}(t) C_m^{(1)}(t).$$

Defining $C_m^{(2)}(t) := -H_{m+1,m}^{(2)} \widehat{\boldsymbol{E}}_m^* \boldsymbol{Y}_m^{(2)}(t)$ leads to a succinct expression for the cospatiality relationship between $\boldsymbol{R}_m^{(2)}(t)$ and $\boldsymbol{V}_{m+1}^{(2)}$,

$$\boldsymbol{R}_m^{(2)}(t) = \boldsymbol{V}_{m+1}^{(2)} C_m^{(2)}(t) C_m^{(1)}(t).$$

Inductively, if we start the $k+1$st cycle with the $m+1$st block basis vector from the previous cycle, i.e., if we take $\boldsymbol{V}_1^{(k+1)} = \boldsymbol{V}_{m+1}^{(k)}$, we then obtain for all $k \geq 1$ and $t \geq 0$ that

$$\boldsymbol{Z}_m^{(k)}(t) = \boldsymbol{\mathcal{V}}_m^{(k+1)} \boldsymbol{Y}_m^{(k+1)}(t) C_m^{(k)}(t) \cdots C_m^{(1)}(t) \quad \text{with} \quad \boldsymbol{Y}_m^{(k)}(t) = \left(\mathcal{H}_m^{(k)} + tI\right)^{-1} \widehat{\boldsymbol{E}}_1, \qquad (4.8)$$

$$\boldsymbol{R}_m^{(k)}(t) = \boldsymbol{V}_{m+1}^{(k)} C_m^{(k)}(t) \cdots C_m^{(1)}(t), \qquad (4.9)$$

where

$$\boldsymbol{R}_m^{(k+1)}(t) = \boldsymbol{R}_m^{(k)}(t) - (A + tI)\boldsymbol{Z}_m^{(k)}(t), \qquad (4.10)$$

with

$$C_m^{(1)}(t) = -H_{m+1,m}^{(1)} \widehat{\boldsymbol{E}}_m^* \boldsymbol{Y}_m^{(1)}(t) B$$

$$C_m^{(j)}(t) = -H_{m+1,m}^{(j)} \widehat{\boldsymbol{E}}_m^* \boldsymbol{Y}_m^{(j)}(t), \quad j = 2, \ldots, k. \qquad (4.11)$$

4.2.2. Integral representation of $B(FOM)^2$ error. We are now ready to state our central result on restarted $B(FOM)^2$. All integrals occurring in the following theorem are assumed to exist.

THEOREM 4.3. *Let f be a Stieltjes function. For $k \geq 1$ and $t \geq 0$ with the matrices $C_m^{(j)}(t) \in \mathbb{S}$ as in (4.11), define the matrix-valued function $\Delta_m^{(k)}(z)$ of the complex variable z as*

$$\Delta_m^{(k)}(z) := \int_0^\infty (z+t)^{-1} C_m^{(k)}(t) \cdots C_m^{(1)}(t)\, \mathrm{d}\mu(t). \tag{4.12}$$

Let

$$\boldsymbol{F}_m^{(1)} := \boldsymbol{\mathcal{V}}_m^{(1)} f\big(\mathcal{H}_m^{(1)}\big) \widehat{\boldsymbol{E}}_1 \boldsymbol{B} = \boldsymbol{\mathcal{V}}_m^{(1)} \int_0^\infty \big(\mathcal{H}_m^{(1)} + tI\big)^{-1} \widehat{\boldsymbol{E}}_1 \boldsymbol{B}\, \mathrm{d}\mu(t)$$

be the $B(FOM)^2$ approximation to $f(A)\boldsymbol{B}$ after the first cycle. For $k \geq 1$ put

$$\widetilde{\boldsymbol{D}}_m^{(k)} := \boldsymbol{\mathcal{V}}_m^{(k+1)} \left(\Delta_m^{(k)}\big(\mathcal{H}_m^{(k+1)}\big) \circ \widehat{\boldsymbol{E}}_1 \right) \tag{4.13}$$

$$:= \boldsymbol{\mathcal{V}}_m^{(k+1)} \int_0^\infty \big(\mathcal{H}_m^{(k+1)} + tI\big)^{-1} \widehat{\boldsymbol{E}}_1 C_m^{(k)}(t) \cdots C_m^{(1)}(t)\, \mathrm{d}\mu(t),$$

$$\boldsymbol{F}_m^{(k+1)} := \boldsymbol{F}_m^{(k)} + \widetilde{\boldsymbol{D}}_m^{(k)}.$$

Then for $k = 0, 1, \ldots$ the kth $B(FOM)^2$ approximation error $\boldsymbol{D}_m^{(k+1)} := f(A)\boldsymbol{B} - \boldsymbol{F}_m^{(k+1)}$ is given as

$$\boldsymbol{D}_m^{(k+1)} = \Delta_m^{(k+1)}(A) \circ \boldsymbol{V}_{m+1}^{(k+1)} := \int_0^\infty (A + tI)^{-1} \boldsymbol{V}_{m+1}^{(k+1)} C_m^{(k+1)}(t) \cdots C_m^{(1)}(t)\, \mathrm{d}\mu(t). \tag{4.14}$$

Before starting the proof, we note that the representations (4.13) and (4.14) are not of the form "matrix function times a block vector," as is the main object of our efforts $f(A)\boldsymbol{B}$. Rather, they are more closely related to the action of a matrix polynomial on a matrix and block vector pair, as defined by the operator $\circ$ in (2.5). By analogy, we have thus extended the meaning of $\circ$ to integrals with matrix-valued coefficients in the definitions of (4.13) and (4.14). With this in mind, one can see that the correction $\widetilde{\boldsymbol{D}}_m^{(k)} = \boldsymbol{\mathcal{V}}_m^{(k+1)} \left(\Delta_m^{(k)}\big(\mathcal{H}_m^{(k+1)}\big) \circ \widehat{\boldsymbol{E}}_1 \right)$ to the approximation $\boldsymbol{F}_m^{(k)}$ is the natural extension of the $B(FOM)^2$ approximation to the matrix function $\Delta_m^{(k)}(A) \circ \boldsymbol{V}_{m+1}^{(k)} = \boldsymbol{D}_m^{(k)}$. in the space $\mathscr{K}_m^{\mathbb{S}}(A, \boldsymbol{V}_1^{(k+1)})$ with $\boldsymbol{V}_1^{(k+1)} = \boldsymbol{V}_{m+1}^{(k)}$.

Proof. of Theorem 4.3. We will make use of the restarted BFOM iterates $\boldsymbol{X}_m^{(k)}(t)$ for the shifted linear systems $(A + tI)\boldsymbol{X}_*(t) = \boldsymbol{B}$, together with their updates $\boldsymbol{Z}_m^{(k)}(t)$, their errors $\boldsymbol{E}_m^{(k)}(t) := \boldsymbol{X}_*(t) - \boldsymbol{X}_m^{(k)}(t)$, and their residuals $\boldsymbol{R}_m^{(k)}(t) = \boldsymbol{B} - (A + tI)\boldsymbol{X}_m^{(k)}(t) = \boldsymbol{V}_{m+1}^{(k)} C_m^{(k)}(t) \cdots C_m^{(1)}(t)$; see equations (4.7), (4.8), and (4.9). Note that, in particular, for all $k = 0, 1, \ldots$, the error representaton for $\boldsymbol{D}_m^{(k+1)}$ from (4.14) is equivalent to

$$\boldsymbol{D}_m^{(k+1)} = \int_0^\infty (A + tI)^{-1} \boldsymbol{R}_m^{(k+1)}(t)\, \mathrm{d}\mu(t).$$

We obtain (4.14) for $k = 0$ via

$$\boldsymbol{D}_m^{(1)} = f(A)\boldsymbol{B} - \boldsymbol{F}_m^{(1)} = \int_0^\infty (A + tI)^{-1}\boldsymbol{B} - \boldsymbol{\mathcal{V}}_m^{(1)}\boldsymbol{Y}_m^{(1)}(t)B\,\mathrm{d}\mu(t)$$

$$= \int_0^\infty (A + tI)^{-1}\boldsymbol{B} - \boldsymbol{X}_m^{(1)}(t)\,\mathrm{d}\mu(t)$$

$$= \int_0^\infty (A + tI)^{-1}\boldsymbol{R}_m^{(1)}(t)\,\mathrm{d}\mu(t).$$

Inductively, then, for $k > 0$ we express $\boldsymbol{D}_m^{(k+1)} = f(A)\boldsymbol{B} - \boldsymbol{F}_m^{(k+1)} = f(A)\boldsymbol{B} - (\boldsymbol{F}_m^{(k)} + \widetilde{\boldsymbol{D}}_m^{(k)}) = \boldsymbol{D}_m^{(k)} - \widetilde{\boldsymbol{D}}_m^{(k)}$ and use (4.8) and (4.9) to obtain that

$$\boldsymbol{D}_m^{(k+1)} = \int_0^\infty (A + tI)^{-1}\boldsymbol{R}_m^{(k)}(t)\,\mathrm{d}\mu(t) - \boldsymbol{\mathcal{V}}_m^{(k+1)}\int_0^\infty \left(\mathcal{H}_m^{(k+1)} + tI\right)^{-1}\widehat{\boldsymbol{E}}_1 C_m^{(k)}(t)\cdots C_m^{(1)}(t)\,\mathrm{d}\mu(t)$$

$$= \int_0^\infty (A + tI)^{-1}\boldsymbol{R}_m^{(k)}(t) - \boldsymbol{Z}_m^{(k)}(t)\,\mathrm{d}\mu(t)$$

$$= \int_0^\infty (A + tI)^{-1}\left(\boldsymbol{R}_m^{(k)}(t) - (A + tI)\boldsymbol{Z}_m^{(k)}(t)\right)\,\mathrm{d}\mu(t)$$

$$= \int_0^\infty (A + tI)^{-1}\boldsymbol{R}_m^{(k+1)}(t)\,\mathrm{d}\mu(t),$$

where the last inequality holds by (4.10). □

Theorem 4.3 provides the restart technique for our proposed B(FOM)^2 algorithm, given as Algorithm 2. Note that the cospatial factors $C_m^{(k)}(t)$ can be retrieved for any t as long $\mathcal{H}_m^{(k)}$ is stored for all k. We employ numerical quadrature to evaluate the integral defining $\widetilde{\boldsymbol{D}}_m^{(k)}$, but one should note that the choice of quadrature rule does not affect the integral representation of the error for the next cycle.

Algorithm 2: B(FOM)^2 with restarts

Given: f, A, $\boldsymbol{B}$, $\mathbb{S}$, $\langle\!\langle\cdot,\cdot\rangle\!\rangle_{\mathbb{S}}$, N, m, `tol`

1 Run Algorithm 1 with inputs A, $\boldsymbol{B}$, $\mathbb{S}$, $\langle\!\langle\cdot,\cdot\rangle\!\rangle_{\mathbb{S}}$, N, and m to obtain $\boldsymbol{\mathcal{V}}_m^{(1)}$, $\mathcal{H}_m^{(1)}$, $H_{m+1,m}^{(1)}$, $\boldsymbol{V}_{m+1}^{(1)}$, and B

2 Compute $\boldsymbol{F}_m^{(1)} = \boldsymbol{\mathcal{V}}_m^{(1)}f\left(\mathcal{H}_m^{(1)}\right)\widehat{\boldsymbol{E}}_1 B$

3 **for** $k = 1, 2, \ldots,$ *until convergence* **do**

4   Determine $C_m^{(k)}(t)$ to define the new error function $\Delta_m^{(k)}(z)$

5   Run Algorithm 1 with inputs A, $\boldsymbol{V}_{m+1}^{(k)}$, $\mathbb{S}$, $\langle\!\langle\cdot,\cdot\rangle\!\rangle_{\mathbb{S}}$, N, and m to obtain $\boldsymbol{\mathcal{V}}_m^{(k+1)}$, $\mathcal{H}_m^{(k+1)}$, $H_{m+1,m}^{(k+1)}$, and $\boldsymbol{V}_{m+1}^{(k+1)}$

6   Compute $\widetilde{\boldsymbol{D}}_m^{(k)} := \boldsymbol{\mathcal{V}}_m^{(k+1)}\Delta_m^{(k)}\left(\mathcal{H}_m^{(k+1)}\right) \circ \widehat{\boldsymbol{E}}_1$, where $\Delta_m^{(k)}(z)$ is evaluated via quadrature

7   Compute $\boldsymbol{F}_m^{(k+1)} := \boldsymbol{F}_m^{(k)} + \widetilde{\boldsymbol{D}}_m^{(k)}$

4.3. Error bounds and convergence results. It remains to investigate when Algorithm 2 produces a convergent sequence of approximations to $f(A)\boldsymbol{B}$. The following lemma is important in this context.

LEMMA 4.4. *Let A be $\langle \cdot, \cdot \rangle_{\mathbb{S}}$-positive-definite, and let $g : (0, \infty) \to (0, \infty)$ be a scalar rational function. Also, let $g_{\min}$ and $g_{\max}$ denote the minimum and maximum values of g on $\mathrm{spec}(A)$, respectively. Then*

$$\sqrt{g_{\min}} \, \|\boldsymbol{V}\|_{\mathbb{S}} \leq \|\boldsymbol{V}\|_{\mathbb{S}\text{-}g(A)} \leq \sqrt{g_{\max}} \, \|\boldsymbol{V}\|_{\mathbb{S}} \, .$$

Proof. Since A is $\langle \cdot, \cdot \rangle_{\mathbb{S}}$-positive definite, its spectrum is positive and A has a $\langle \cdot, \cdot \rangle_{\mathbb{S}}$-orthonormal eigenbasis, i.e., there exist $\{\beta_j\}_{j=1}^{ns} \subset \mathbb{C}$ and $\{\boldsymbol{Q}_j\}_{j=1}^{ns} \subset \mathbb{C}^{n \times s}$ such that $A\boldsymbol{Q}_j = \lambda_j \boldsymbol{Q}_j$ and $\langle \boldsymbol{Q}_j, \boldsymbol{Q}_k \rangle_{\mathbb{S}} = \delta_{jk}$. Given any $\boldsymbol{V} \in \mathbb{C}^{n \times s}$, expand it in terms of this basis as $\boldsymbol{V} = \sum_{j=1}^{ns} \beta_j \boldsymbol{Q}_j$. Then

$$\|\boldsymbol{V}\|_{\mathbb{S}\text{-}g(A)}^2 = \langle g(A)\boldsymbol{V}, \boldsymbol{V} \rangle_{\mathbb{S}} = \langle \sum_{j=1}^{ns} g(\lambda_j)\beta_j \boldsymbol{Q}_j, \sum_{j=1}^{ns} \beta_j \boldsymbol{Q}_j \rangle_{\mathbb{S}} = \sum_{j=1}^{ns} g(\lambda_j)|\beta_j|^2,$$

and thus

$$g_{\min} \sum_{j=1}^{ns} |\beta_j|^2 \leq \|\boldsymbol{V}\|_{\mathbb{S}\text{-}g(A)}^2 \leq g_{\max} \sum_{j=1}^{ns} |\beta_j|^2.$$

Noting that $\sum_{j=1}^{ns} |\beta_j|^2 = \|\boldsymbol{V}\|_{\mathbb{S}}^2$ leads to the desired result. $\square$

Define the following shifted versions of (3.6):

$$\kappa(t) := \frac{\lambda_{\max} + t}{\lambda_{\min} + t}, \quad c(t) := \frac{\sqrt{\kappa(t)} - 1}{\sqrt{\kappa(t)} + 1}, \quad \text{and} \quad \xi_m(t) := \frac{1}{\cosh(m \ln c(t))}. \tag{4.15}$$

Note that for all $t \geq 0$, $0 \leq \xi_m(t) < 1$, and $\lim_{t \to \infty} \xi_m(t) = 0$; see [24, Proposition 4.2].

The following theorem generalizes the results of [24, Lemma 4.1 and Theorem 4.3] to the block case. Incidentally, its proof also shows that the improper integral representation (4.14) for the error is finite under the given assumptions on A, which at the same time implies that the restarted B(FOM)^2 approximations are all well-defined.

THEOREM 4.5. *Let f be a Stieltjes function, $A \in \mathbb{C}^{n \times n}$ $\langle\!\langle \cdot, \cdot \rangle\!\rangle_{\mathbb{S}}$-self-adjoint and $\langle \cdot, \cdot \rangle_{\mathbb{S}}$-positive definite, and $\boldsymbol{B} \in \mathbb{C}^{n \times s}$. Let $\boldsymbol{D}_m^{(k)}$ from (4.14) be the error of the restarted B(FOM)^2 approximation after k cycles of length m. Then, with the quantities defined in (4.15), we have*

$$\left\| \boldsymbol{D}_m^{(k)} \right\|_{\mathbb{S}\text{-}A} \leq \|\boldsymbol{B}\|_{\mathbb{S}} \sqrt{\lambda_{\max}} \int_0^\infty \frac{\xi_m(t)^k}{\sqrt{\lambda_{\min} + t}\sqrt{\lambda_{\max} + t}} \, \mathrm{d}\mu(t) \leq \gamma \xi_m(0)^k, \tag{4.16}$$

where $\gamma = \|\boldsymbol{B}\|_{\mathbb{S}} \sqrt{\lambda_{\max}} f(\sqrt{\lambda_{\min}\lambda_{\max}})$. In particular, B(FOM)^2 converges for all cycle lengths m.

Proof. We begin by writing the error $\boldsymbol{D}_m^{(k)}$ as an integral over the shifted restarted BFOM error as

$$\boldsymbol{D}_m^{(k)} = \int_0^\infty (A + tI)^{-1} \boldsymbol{R}_m^{(k)} \, \mathrm{d}\mu(t) = \int_0^\infty \boldsymbol{E}_m^{(k)}(t) \, \mathrm{d}\mu(t).$$

The norm can be passed through the integral (see, e.g., [56, Lemma 2.1]) to give

$$\left\| \boldsymbol{D}_m^{(k)} \right\|_{\mathbb{S}\text{-}A} \leq \int_0^\infty \left\| \boldsymbol{E}_m^{(k)}(t) \right\|_{\mathbb{S}\text{-}A} \, \mathrm{d}\mu(t).$$

Using Lemma 4.4, we have for any $\boldsymbol{V} \in \mathbb{C}^{n \times s}$ that

$$\|\boldsymbol{V}\|_{\mathbb{S}\text{-}A+tI}^2 = \langle \boldsymbol{V}, \boldsymbol{V}\rangle_{\mathbb{S}\text{-}A} + t\langle \boldsymbol{V}, \boldsymbol{V}\rangle_{\mathbb{S}} = \|\boldsymbol{V}\|_{\mathbb{S}\text{-}A}^2 + t\|\boldsymbol{V}\|_{\mathbb{S}}^2 \geq \|\boldsymbol{V}\|_{\mathbb{S}\text{-}A}^2 + \tfrac{t}{\lambda_{\max}} \|\boldsymbol{V}\|_{\mathbb{S}\text{-}A} = \tfrac{\lambda_{\max}+t}{\lambda_{\max}} \|\boldsymbol{V}\|_{\mathbb{S}\text{-}A}^2 ,$$

implying that $\|\boldsymbol{V}\|_{\mathbb{S}\text{-}A} \leq \sqrt{\tfrac{\lambda_{\max}}{\lambda_{\max}+t}} \|\boldsymbol{V}\|_{\mathbb{S}\text{-}A+tI}$. Then

$$\left\| \boldsymbol{E}_m^{(k)}(t) \right\|_{\mathbb{S}\text{-}A} \leq \sqrt{\frac{\lambda_{\max}}{\lambda_{\max}+t}} \left\| \boldsymbol{E}_m^{(k)}(t) \right\|_{\mathbb{S}\text{-}A+tI}. \tag{4.17}$$

Repeated application of Theorem 3.7 to $\left\| \boldsymbol{E}_m^{(k)}(t) \right\|_{\mathbb{S}\text{-}A+tI}$ gives that

$$\left\| \boldsymbol{E}_m^{(k)}(t) \right\|_{\mathbb{S}\text{-}A+tI} \leq \xi_m(t)^k \left\| \boldsymbol{E}_0^{(1)}(t) \right\|_{\mathbb{S}\text{-}A+tI}. \tag{4.18}$$

Since $\boldsymbol{E}_0^{(1)}(t) = \boldsymbol{X}_*(t)$, we can use Lemma 4.4 to bound

$$\left\| \boldsymbol{E}_0^{(1)}(t) \right\|_{\mathbb{S}\text{-}A+tI}^2 = \langle (A+tI)^{-1}\boldsymbol{B}, (A+tI)(A+tI)^{-1}\boldsymbol{B}\rangle_{\mathbb{S}} = \|\boldsymbol{B}\|_{\mathbb{S}\text{-}(A+tI)^{-1}}^2 \leq \frac{1}{\lambda_{\min}+t} \|\boldsymbol{B}\|_{\mathbb{S}}^2 .$$

Combining (4.17), (4.18), (4.19) we obtain that

$$\left\| \boldsymbol{D}_m^{(k)} \right\|_{\mathbb{S}\text{-}A} \leq \|\boldsymbol{B}\|_{\mathbb{S}} \sqrt{\lambda_{\max}} \int_0^\infty \frac{\xi_m(t)^k}{\sqrt{\lambda_{\min}+t}\sqrt{\lambda_{\max}+t}} \, \mathrm{d}\mu(t), \tag{4.19}$$

which is the first inequality in (4.16). The bound to the right of (4.19) is increased if we replace $\xi_m(t)$ by $\xi_m(0)$, since $\xi_m(t)$ is a monotonically decreasing function of t [24, Proposition 4.2]. Moreover, the geometric mean $\sqrt{\lambda_{\min}\lambda_{\max}}$ satisfies

$$\frac{1}{\sqrt{\lambda_{\min}+t}\sqrt{\lambda_{\max}+t}} \leq \frac{1}{\sqrt{\lambda_{\min}\lambda_{\max}}+t},$$

so that

$$\int_0^\infty \frac{\xi_m(t)^k}{\sqrt{\lambda_{\min}+t}\sqrt{\lambda_{\max}+t}} \, \mathrm{d}\mu(t) \leq \xi_m(0)^k \int_0^\infty \frac{1}{\sqrt{\lambda_{\min}\lambda_{\max}}+t} \, \mathrm{d}\mu(t).$$

Observing that the integral on the right is just $f(\sqrt{\lambda_{\min}\lambda_{\max}})$ leads to the second inequality in (4.16). $\square$

5. Numerical experiments. In this section, we illustrate the behavior of restarted B(FOM)^2 for a variety of functions—not all of which are Stieltjes—and matrices—not all of which are $\langle\!\langle \cdot, \cdot \rangle\!\rangle_{\mathbb{S}}$-self-adjoint or $\langle \cdot, \cdot \rangle_{\mathbb{S}}$-positive definite. While timings and computational effort, especially in comparison to the non-block method, are important in practice, we do not devote much attention to them here, as the main purpose of these examples is to establish the versatility and applicability of B(FOM)^2. Four versions are implemented, ClB(FOM)^2, GlB(FOM)^2, LiB(FOM)^2, and HyB(FOM)^2 (which takes $\langle\!\langle \cdot, \cdot \rangle\!\rangle_{\mathbb{S}}^{\mathrm{Hy}}$ as in (2.1) and $N^{\mathrm{Hy}}(\boldsymbol{X}) := \mathtt{sqrtm}(\boldsymbol{X}^*\boldsymbol{X})$), along with the non-block approach of [24] applied to each column in serial, referred to as (FOM)^2. For all four corresponding block inner products, $A = A^H$ implies that A is $\langle\!\langle \cdot, \cdot \rangle\!\rangle_{\mathbb{S}}$-self-adjoint.

5.1. Expressions for the matrix error function. We consider three functions, $z^{-\alpha}$, $\frac{\log(z+1)}{z}$, and $\exp(z)$. For the respective integral representation of the error functions we use the quadrature rules detailed in [25], which we now describe shortly.

5.1.1. $f(z) = z^{-\alpha}$, $0 < \alpha < 1$. Applying the Cayley transform $t = \delta\frac{1-x}{1+x}$, for some $\delta > 0$, to (4.5), we use N-node Gauss-Jacobi quadrature for the interval $[-1, 1]$ (as in, e.g., [13]) and find

$$\widetilde{\boldsymbol{D}}_m^{(k)} \approx -c_{\alpha,\delta} \sum_{j=1}^{N} \frac{w_j}{1+x_j} \boldsymbol{\mathcal{V}}_m^{(k+1)}(\mathcal{H}_m^{(k+1)} + t_j I)^{-1}\widehat{\boldsymbol{E}}_1 C_m^{(k)}(t_j) \ldots C_m^{(1)}(t_j),$$

with the Gauss-Jacobi nodes $\{x_j\}_{j=1}^{N}$, weights $\{w_j\}_{j=1}^{N}$, and $\{t_j := \delta\frac{1-x_j}{1+x_j}\}_{j=1}^{N}$. As discussed in [25], the algorithm is not sensitive to the choice of δ, so we take $\delta = 1$.

5.1.2. $f(z) = \log(z + 1)/z$. With the Gauss-Legendre nodes $\{x_j\}_{j=1}^{N}$, weights $\{w_j\}_{j=1}^{N}$, and transformed nodes $\{t_j := \frac{2}{1-x_j}\}_{j=1}^{N}$, we obtain

$$\widetilde{\boldsymbol{D}}_m^{(k)} \approx \sum_{j=1}^{N} \frac{w_j}{1-x_j} \boldsymbol{\mathcal{V}}_m^{(k+1)}(\mathcal{H}_m^{(k+1)} + t_j I)^{-1}\widehat{\boldsymbol{E}}_1 C_m^{(k)}(t_j) \cdots C_m^{(1)}(t_j).$$

5.1.3. $f(z) = \exp(z)$. Although the exponential is not a Stieltjes function, we can still apply the framework developed in this paper to the Cauchy integral form,

$$\exp(z) = \frac{1}{2\pi i} \int_{\Gamma} \frac{\exp(t)}{t-z} \, dt. \tag{5.1}$$

Based on work in [63, 64, 65], the authors of [25] recommend taking Γ as a parabolic contour parametrized as

$$\gamma(s) = a + is - cs^2, \quad s \in \mathbb{R}.$$

Here, a and c are chosen anew for each restart cycle to ensure that Γ encloses the eigenvalues of the matrix $\mathcal{H}_m$. The infinite interval of integration for s is truncated for a given error tolerance `tol` by the truncation parameter $s_0 := \sqrt{a - \log(\texttt{tol})/c}$, so that $|\exp(\gamma(\pm s_0))| = \texttt{tol}$. From the N-point midpoint rule on $[-s_0, s_0]$ we obtain the nodes $s_j := s_0(\frac{2j-1}{N})$, $j = 1, \ldots, N$. Defining $w_j := \exp(\gamma(s_j))\gamma'(s_j)$ and $t_j := -\gamma(s_j)$, we then approximate the error approximation as

$$\widetilde{\boldsymbol{D}}_m^{(k)} \approx \frac{s_0}{N\pi i} \sum_{j=1}^{N} w_j \boldsymbol{\mathcal{V}}_m^{(k+1)}(\mathcal{H}_m^{(k+1)} + t_j I)^{-1}\widehat{\boldsymbol{E}}_1 C_m^{(k)}(t_j) \cdots C_m^{(1)}(t_j).$$

5.2. Remarks on implementation. We highlight some aspects relevant to an implementation of the algorithms described in this paper.

The first aspect pertains to breakdowns of the block Arnoldi process due to the occurrence of a rank-deficient block vector $\boldsymbol{W}$ in Algorithm 1. For the global method, this situation implies that $\boldsymbol{W} = 0$, i.e., that $\mathscr{K}_m^{\text{Gl}}$ has reached its maximal size, the algorithm has converged, and $f(A)\boldsymbol{B} = \boldsymbol{F}_m^{(k)}$. Likewise, for the loop interchanged method, an zero entry in the i-th diagonal position of $\langle\!\langle \boldsymbol{W}, \boldsymbol{W} \rangle\!\rangle_{\mathbb{S}}^{\text{Li}}$ indicates that the process has converged for the ith column of $\boldsymbol{B}$, so this column can be discarded and the process can continue with the remaining ones.

The classical method, however, faces a more complicated issue. Algorithm 1 then may generate singular block entries $H_{k+1,k}$ even when none of the columns has converged, due to possible linear dependence amongst columns of the basis vector $\boldsymbol{V}_k$. A number of methods have been proposed to manage this phenomenon; see, e.g., [8, 33] and references therein. We employ [8, Algorithm 7.3], which uses a rank-revealing QR (RRQR) factorization of $\boldsymbol{W}$ to detect near and exact linear dependence. The RRQR acts as the scaling quotient N^{Cl} instead of R from the standard economic QR factorization. We demonstrate the behavior of a RRQR-based implementation of the classical method versus the standard implementation in section 5.4.

The other key aspect is that we use quadrature to evaluate the error function (4.12), which directly affects the quality of the solution update $\widetilde{\boldsymbol{D}}_m^{(k)}$. Specifically, we implement adaptive quadrature, in which the number of nodes is increased by $\sqrt{2}$ until the difference between two consecutive applications of the quadrature rule is less than a given tolerance. The final number of nodes is stored, and in the next cycle, slightly fewer nodes are first used in an attempt to decrease the computational effort; if they are not sufficient, the nodes are increased again until the tolerance is met, and so on. For all examples in this paper, the quadrature tolerance is set to be the same as the convergence tolerance.

As we have access to the exact solution $f(A)\boldsymbol{B}$ (or, rather, a machine-accurate approximation to it) for each example in this paper, we use $\boldsymbol{D}_m^{(k)}$, i.e., the difference between the approximation $\boldsymbol{F}_m^{(k)}$ and $f(A)\boldsymbol{B}$, to measure convergence. A process is regarded as convergent when $\left\|\boldsymbol{D}_m^{(k)}\right\|_{A\text{-F}}$ is below a given error tolerance, which is specified for each example.

All experiments are run on a Dell desktop with a Linux 64-bit operating system, an Intel®Core$^{\text{TM}}$ i7-4770 CPU @ 3.40 GHz, and 32 GB of RAM. In the plots, we abbreviate ClB(FOM)^2, GlB(FOM)^2, LiB(FOM)^2, and (FOM)^2 as Cl, Gl, Li, and nB, respectively, where "nB" stands for "non-block."

5.3. B(FOM)^2 on a random tridiagonal HPD matrix. Our first example consists of the function $f(z) = z^{-1/2}$ acting on a 100×100 HPD tridiagonal matrix A generated from random entries with a condition number of $O(10^2)$. The cycle length m is set to 5, and the error tolerance is set to 1e-10. For the hybrid method, $q = 5$; i.e., $\mathbb{S}^{\text{Hy}}$ consists of 10×10 matrices with 5×5 blocks on the diagonal. With this example we illustrate that while the error bound given in Theorem 4.5 is valid, it is too far from the actual values of the error to be used as a predictor of convergence. Compare the black line at 10^1 with the rest of the convergence curves in Figure 5.1. (Note that GlB(FOM)^2, LiB(FOM)^2, and (FOM)^2 all have nearly identical error curves and therefore appear to overlap each other. In fact, LiB(FOM)^2 and (FOM)^2 should be identical in exact arithmetic.) Since $\xi_m(0)$ is very close to 1, this bound cannot precisely predict the convergence exhibited by the actual error curves. We therefore do not include it in further convergence plots.

5.4. Discretized two-dimensional Laplacian and $f(z) = z^{-1/2}$. We now take A to be the real, symmetric positive definite matrix arising from the second-order central difference discretization of the negative two-dimensional Laplacian operator with 100 grid points in each spatial dimension, so that $n = 10^4$. We look at two different $\boldsymbol{B}$ with $s = 10$, the first having full rank and the second being rank deficient. We also run two versions of ClB(FOM)^2, one with deflation and one without, as described in section 5.2, in order to observe the effects of rank deficiency on the stability of ClB(FOM)^2. For both cases, the cycle length is $m = 25$, and the error tolerance is 1e-6. HyB(FOM)^2 is not run for this set of examples, due to the complexity of developing an adequate deflation procedure. We take $\boldsymbol{B} = \boldsymbol{u} \otimes I_{10}$, where $\boldsymbol{u}$ is the vector of dimension 10^3 whose entries are all 1, and I_{10} is the 10×10 identity, making $\boldsymbol{B}$ full rank.
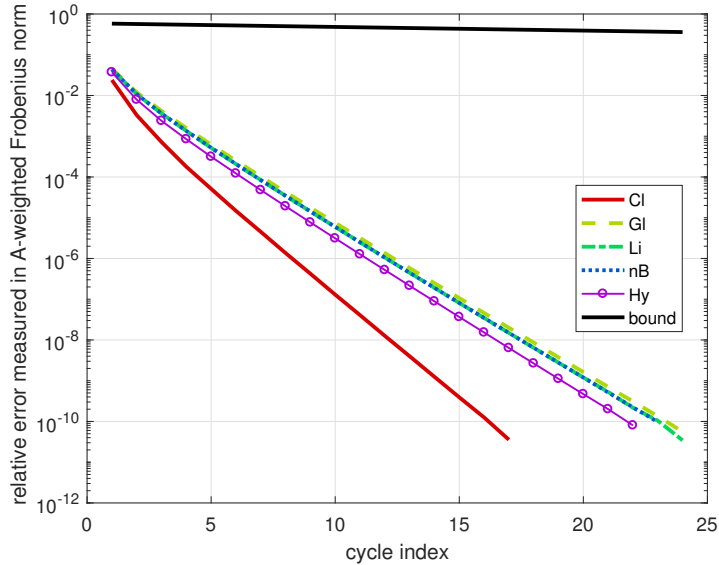
Fig. 5.1: Convergence history for computing $A^{-1/2}\boldsymbol{B}$, where $A \in \mathbb{C}^{100 \times 100}$ is a random tridiagonal HPD matrix, and $\boldsymbol{B} \in \mathbb{C}^{100 \times 10}$ is random.

The left part of Figure 5.2 displays the results for $\boldsymbol{B}$ having full rank. One can see that all methods attain the desired accuracy with roughly the same number of cycles. The error curves for ClB(FOM)^2 with and without deflation coincide, as well as those of LiB(FOM)^2, $(\text{FOM})^2$, and GlB(FOM)^2, leading to what may appear to be only two curves. However, for a given cycle, GlB(FOM)^2 is slightly less accurate than LiB(FOM)^2 or $(\text{FOM})^2$, and all are less accurate than ClB(FOM)^2, as predicted by the comparative result (3.13).

For the right part of Figure 5.2, the first column of $\boldsymbol{B}$ is a linear combination of the four other columns. The right figure shows that ClB(FOM)^2 stagnates almost immediately, because it does not react to the rank deficiency. Indeed, when computing R in the economic QR-factorization of $\boldsymbol{B}$ as a scaling quotient, the algorithm treats the R factor as full rank, since the smallest element on the diagonal is of $O(10^{-11})$, i.e., neither exactly zero nor numerically singular. Taking inverses introduces enough numerical error so that by the next cycle, the basis vector (now inaccurate) has full rank. It halts once the code detects that the norm of the error is no longer monotonically decreasing. In contrast, ClB(FOM)^2 with deflation converges correctly, albeit with slow runtime. This is due to the comparatively high cost for the RRQR factorization. Although the cost per cycle is still high compared to, e.g., GlB(FOM)^2 and $(\text{FOM})^2$, significantly fewer cycles are required overall.

It is important to note that this situation is contrived, and linear dependence amongst the columns of basis vectors is unlikely to occur in most practical situations. However, given how slow the runtime of ClB(FOM)^2 with deflation is in our Matlab implementation – roughly 20 times slower in the full-rank example – we do not examine it in further examples. Furthermore, given the excellent performance of GlB(FOM)^2, which requires no special treatment of rank deficiencies,
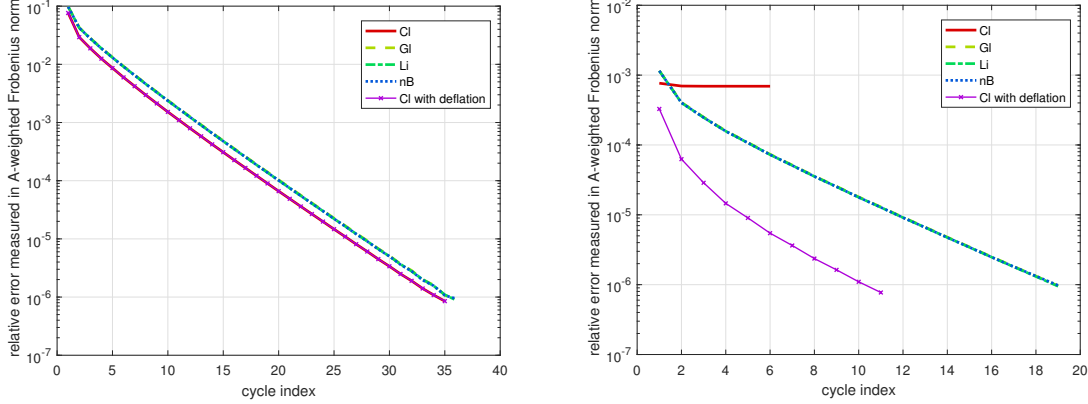
Fig. 5.2: Convergence history for $A^{-1/2}\boldsymbol{B}$, where $A \in \mathbb{C}^{10^4 \times 10^4}$ is the discretized two-dimensional Laplacian. Left: $\boldsymbol{B} \in \mathbb{C}^{10^4 \times 10}$ has linearly independent columns. Right: the first column of $\boldsymbol{B}$ is a linear combination of other columns.

tailoring ClB(FOM)^2 to work in rank deficient cases may not be necessary.

5.5. Overlap Dirac operator and $f(z) = \text{sign}(z)$**.** Quantum chromodynamics (QCD) is an area of theoretical particle physics in which the strong interaction between quarks and gluons is studied. Simulations are carried out on a four-dimensional space-time lattice, where each lattice point is assigned 12 variables, and each variable corresponds to a combination of three colors and four spins. One must solve systems involving the overlap Dirac operator, described, e.g., in [23, 46]. The solution of such systems entails the computation of $\text{sign}(Q)\widehat{\boldsymbol{E}}_1$, where the matrix Q represents a periodic nearest-neighbor coupling on the lattice, and $\widehat{\boldsymbol{E}}_1$ is the first 12 unit vectors of the lattice space. The matrix Q is large, sparse, complex and Hermitian if the chemical potential is zero, which is what we consider here.

We take an 8^4 lattice, such that $n = 12 \cdot 8^4 = 49152$, $A = Q^2 \in \mathbb{C}^{49152 \times 49152}$. Using $\text{sign}(z) = (z^2)^{-1/2}z$, we compute $\text{sign}(Q)\widehat{\boldsymbol{E}}_1$ as $A^{-1/2}\boldsymbol{B}$, where $\boldsymbol{B} = Q\widehat{\boldsymbol{E}}_1 \in \mathbb{C}^{49152 \times 12}$. We set the error tolerance to 1e-6 and vary the cycle length, letting $m \in \{25, 50, 100, 150\}$. We also include HyB(FOM)^2 in this series of tests, with $q = 4$.

Figure 5.3 shows the correlation between the cycle length and the number of cycles required to converge. Aside from GlB(FOM)^2 requiring one extra cycle when $m = 50$, all methods require the same number of cycles to converge for a given m. In which case, a computationally less intensive method, such as LiB(FOM)^2 or GlB(FOM)^2, should be preferred, since ClB(FOM)^2 does not take advantage of the extra information in its larger approximation space as m increases.

5.6. Bus power system and $\log(z + 1)/z$**.** We now consider the Stieltjes function $f(z) = \log(z+1)/z$, and its action on the S-admittance matrix A of a bus power system, specifically matrix HB/1138_bus of the University of Florida Sparse Matrix Collection [14]. The symmetric matrix A has a condition number of $\mathcal{O}(10^6)$, and its sparsity pattern is given in the left of Figure 5.4. The block vector $\boldsymbol{B}$ has random entries, and we vary the number of its columns, i.e., $s \in \{2, 6, 12, 24\}$, while the cycle length m is kept constant at $m = 50$. We include HyB(FOM)^2 in this series of tests
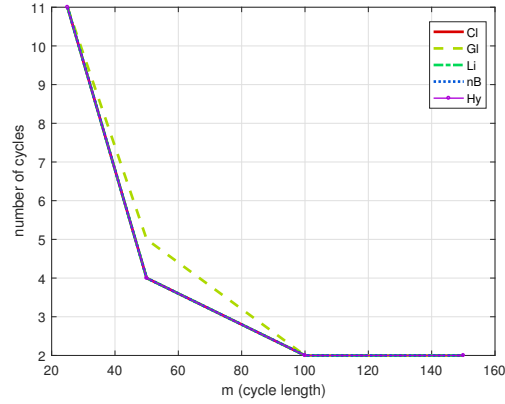
Fig. 5.3: Number of cycles versus the cycle length for the overlap Dirac operator example.

as well, with $q \in \{1, 2, 4, 6\}$. Note that with $q = 1$, HyB(FOM)^2 is arithmetically equivalent to LiB(FOM)^2.

As shown in Figure 5.4 LiB(FOM)^2 and (FOM)^2 require roughly the same number of cycles to converge, with GlB(FOM)^2 requiring the next largest number of cycles. For both ClB(FOM)^2 and HyB(FOM)^2, the cycle count decreases as s increases. Since $\mathbb{S}^{Cl} = \mathbb{C}^{s \times s}$, the classical method obtains its iterates from the largest possible block Krylov space among the methods. In particular, it takes information from $m \cdot s = 50s$ columns, which, as s increases, becomes quite large relative to the problem size $n = 1138$, hence why ClB(FOM)^2 benefits so substantially from the additional columns in this example. Considering that HyB(FOM)^2 is a mixture of LiB(FOM)^2 and ClB(FOM)^2, it is
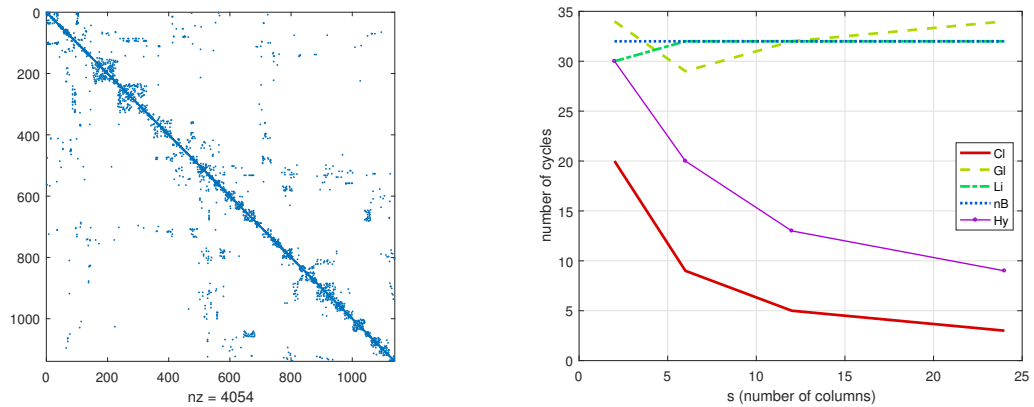

Fig. 5.4: Left: Sparsity pattern of the bus power system matrix. Right: Number of cycles needed for convergence versus s, the number of columns.

quite noteworthy that the method exhibits trends closer to those of ClB(FOM)^2 than of LiB(FOM)^2.

5.7. Convection-diffusion equation and $\exp(z)$**.** In the final example, we look at the action of the exponential, on a series of matrices, which vary in degree of non-symmetry. The matrices correspond to the standard finite differences discretization of a two-dimensional convection-diffusion equation on $[0,1] \times [0,1]$ with a constant convection field and convection parameter ν. We use 350 discretization points in each dimension and a scaling parameter of $2 \cdot 10^{-3}$, resulting in matrices of dimensions $350^2 \times 350^2 = 122,500 \times 122,500$. We look at three matrices A_ν, for $\nu \in \{0, 100, 200\}$. When $\nu = 0$, A_ν is real symmetric; otherwise, A_ν is non-symmetric.

B(FOM)^2 converges in every scenario, as exhibited by the plots in Figure 5.5. In each figure, the curves for ClB(FOM)^2 and HyB(FOM)^2 overlap with each other, and likewise the curves for GlB(FOM)^2, LiB(FOM)^2, and (FOM)^2, resulting in what appears to be only two curves. For a given ν, each method requires the same number of cycles: 4 for $\nu = 0$, 6 for $\nu = 100$, and 9 for $\nu = 200$.
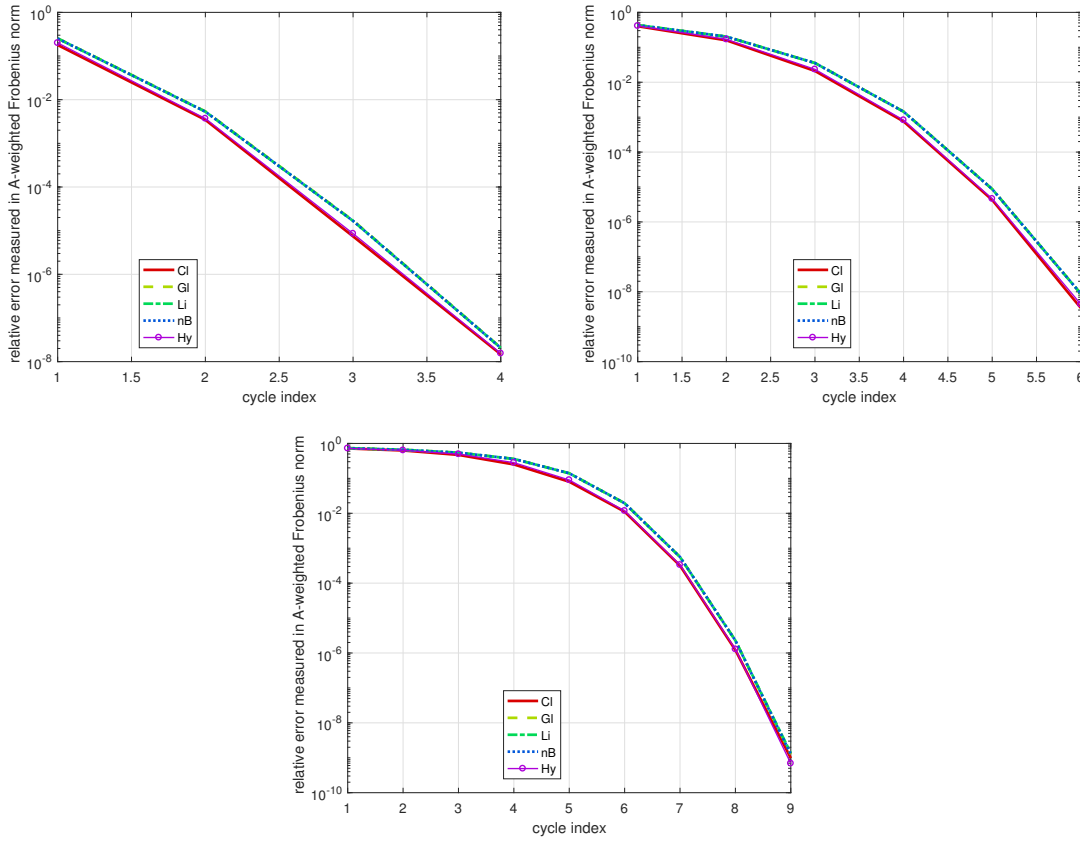

Fig. 5.5: Convergence histories for computing $\exp(A_\nu)\boldsymbol{B}$, where $A_\nu \in \mathbb{C}^{122,500 \times 122,500}$ is the finite differences stencil of a two-dimensional convection-diffusion equation with varying convection parameters $\nu \in \{0, 100, 200\}$, and $\boldsymbol{B} \in \mathbb{C}^{122,500 \times 10}$ has random entries.

6. Conclusions. As matrix functions are featured more in scientific computing, it is vital that a solid theoretical foundation be laid for the methods used to evaluate them. In this work, we have built such a foundation for block Krylov methods for matrix functions within a framework that generalizes existing theory for block Krylov methods for linear systems. This framework, which hinges on a *-subalgebra $\mathbb{S}$, a block inner product $\langle\!\langle \cdot, \cdot \rangle\!\rangle_{\mathbb{S}}$, and a scaling quotient N, opens up avenues for new block Krylov methods, as demonstrated by the success of our hybrid method in the numerical examples.

We have also established a variational characterization of and convergence theory for $\mathrm{B(FOM)}^2$ in the case of $\langle\!\langle \cdot, \cdot \rangle\!\rangle_{\mathbb{S}}$-self-adjoint and $\langle \cdot, \cdot \rangle_{\mathbb{S}}$-positive definite matrices and Stieltjes functions, drawing heavily on conjugate gradients theory. While the resulting convergence bounds are shown to be too pessimistic in subsection 5.3, they comprise a necessary theoretical foundation for $\mathrm{B(FOM)}^2$. Additionally, in our formulation of $\mathrm{B(FOM)}^2$, not only do we account for restarts, but we do so in an efficient way via the matrix error function $\varDelta_m^{(k)}$ and additive correction $\widetilde{\boldsymbol{D}}_m^{(k)}$. The variety of numerical experiments we have presented show that updating with $\widetilde{\boldsymbol{D}}_m^{(k)}$ leads to stable approximations $\boldsymbol{F}_m^{(k)}$.

7. Acknowledgments. The authors would like to thank Tyrone Rees for a template used to make Figure 2.1; Marcel Schweitzer for code that formed the foundation of the implementation of Algorithm 2, data used in section 5.5, and many useful discussions; and Stanley Eisenstat for a copy of [20], which provided insight for the proof of Theorem 3.7.

<div align="center">REFERENCES</div>

[1] Martin Afanasjew, Michael Eiermann, Oliver G. Ernst, and Stefan Güttel, *Implementation of a restarted Krylov subspace method for the evaluation of matrix functions*, Linear Algebra and its Applications, 429 (2008), pp. 229–314.

[2] Awad H. Al-Mohy and Nicholas Higham, *Computing the action of the matrix exponential, with an application to exponential integrators*, SIAM Journal on Scientific Computing, 33 (2010), pp. 488–511.

[3] Fatemeh Panjeh Ali Beik and Davod Khojasteh Salkuyeh, *On the global Krylov subspace methods for solving general coupled matrix equations*, Computers and Mathematics with Applications, 62 (2011), pp. 4605–4613.

[4] José I. Aliaga, Daniel L. Boley, Roland W. Freund, and Vicente Hernández, *A Lanczos-type method for multiple starting vectors*, Mathematics of Computation, 69 (2000), pp. 1577–1601.

[5] Allison H. Baker, John M. Dennis, and Elizabeth R. Jessup, *On improving linear solver performance: a block variant of GMRES*, SIAM Journal on Scientific Computing, 27 (2006), pp. 1608–1626.

[6] Peter Benner, Patrick Kürschner, and Jens Saak, *Frequency-limited balanced truncation with low-rank approximations*, Tech. Report MPIMD/15-09, Max Planck Institute for Dynamics of Complex Technical Systems, July 2015.

[7] Michele Benzi and Valeria Simoncini, *Approximation of functions of large matrices with Kronecker structure*, 2017.

[8] Sebastian Birk, *Deflated shifted block Krylov subspace methods for hermitian positive definite matrices*, Ph.D. Thesis, Fachbereich Mathematik und Naturwissenschaften der Bergischen Universität Wuppertal, 2015.

[9] Rkia Bouyouli, Khalide Jbilou, Rachid Sadaka, and Hassane Sadok, *Convergence properties of some block Krylov subspace methods for multiple linear systems*, Journal of Computational and Applied Mathematics, 196 (2006), pp. 498–511.

[10] Charles George Broyden, *A breakdown of the block CG method*, Optimization Methods and Software, 7 (1997), pp. 41–55.

[11] David Damanik, Alexander Pushnitski, and Barry Simon, *The analytic theory of matrix orthogonal polynomials*, Surveys in Approximation Theory, 4 (2008), pp. 1–85.

[12] Philip I. Davies and Nicholas J. Higham, *Computing $f(A)b$ for matrix functions f*, in QCD and Numerical Analysis III, Artan Boriçi, Andreas Frommer, Bálint Joó, Anthony Kennedy, and Brian Pendelton, eds., vol. 47 of Lecture Notes in Computational Science and Engineering, Berlin, 2005, Springer, pp. 15–24.

[13] Philip J. Davis and Philip Rabinowitz, *Methods of Numerical Integration*, Academic Press, New York, 1975.

[14] Timothy A. Davis and Yifan Hu, *The University of Florida Sparse Matrix Collection*, ACM Transactions on Mathematical Software, 28 (2011), pp. 1:1–1:25.

[15] John E. Dennis, Joseph F. Traub, and R. P. Weber, *The algebraic theory of matrix polynomials*, SIAM Journal on Numerical Analysis, 13 (1976), pp. 831–845.

[16] Richard Dorrance, Fengbo Ren, and Dejan Marković, *A scalable sparse matrix-vector multiplication kernel for energy-efficient sparse-BLAS on FPGAs*, in Proceedings of the 2014 ACM/SIGDA International Symposium on Field-programmable Gate Arrays, FPGA '14, New York, NY, USA, 2014, ACM, pp. 161–170.

[17] Augustin A. Dubrulle, *Retooling the method of block conjugate gradients*, Electronic Transations on Numerical Analysis, 12 (2001), pp. 216–233.

[18] Ian S. Duff, Michael A. Heroux, and Roldan Pozo, *An overview of the sparse basic linear algebra subprograms: The new standard from the BLAS technical forum*, ACM Transactions on Mathematical Software, 28 (2002), pp. 239–267.

[19] Michael Eiermann and Oliver G. Ernst, *A restarted Krylov subspace method for the evaluation of matrix functions*, SIAM Journal on Numerical Analysis, 44 (2006), pp. 2481–2504.

[20] Stanley C. Eisenstat, *On the rate of convergence of B-CG and BGMRES*, 2015. Unpublished.

[21] Ahmed El Guennouni, Khalide Jbilou, and Hassane Sadok, *The block Lanczos method for linear systems with multiple right-hand sides*, Applied Numerical Mathematics, 51 (2004), pp. 243–256.

[22] Lakhdar Elbouyahyaoui, Abderrahim Messaoudi, and Hassane Sadok, *Algebraic properties of the block GMRES and block Arnoldi methods*, Electronic Transactions on Numerical Analysis, 33 (2009), pp. 207–220.

[23] Jasper van den Eshof, Andreas Frommer, Thomas Lippert, Klaus Schilling, and Henk A. van der Vorst, *Numerical methods for the QCD overlap operator. I. Sign-function and error bounds*, Computer Physics Communications, 146 (2002), pp. 203–224.

[24] Andreas Frommer, Stefan Güttel, and Marcel Schweitzer, *Convergence of restarted Krylov subspace methods for Stieltjes functions of matrices*, SIAM Journal on Matrix Analysis and Applications, 35 (2014), pp. 1602–1624.

[25] ——, *Efficient and stable Arnoldi restarts for matrix functions based on quadrature*, SIAM Journal on Matrix Analysis and Applications, 35 (2014), pp. 661–683.

[26] Andreas Frommer, Kathryn Lund, Marcel Schweitzer, and Daniel B. Szyld, *The Radau-Lanczos method for matrix functions*, Tech. Report 16-04-25, Temple University, April 2016. Revised October 2016.

[27] Andreas Frommer and Valeria Simoncini, *Matrix functions*, in Model Order Reduction: Theory, Research Aspects and Applications, Wilhelmus H. A. Schilders, Henk A. van der Vorst, and Joost Rommes, eds., vol. 13 of Mathematics in Industry, Berlin, 2008, Springer, pp. 275–304.

[28] André Gaul, *Recycling Krylov subspace methods for sequences of linear systems– analysis and applications*, Ph.D. Thesis, Technische Universitt Berlin, 2014.

[29] Israel Gohberg, Peter Lancaster, and Leiba Rodman, *Matrix Polynomials*, SIAM, Philadelphia, 2nd ed., 2009.

[30] Chen Greif, Tyrone Rees, and Daniel B. Szyld, *GMRES with multiple preconditioners*, SeMA Journal, (2016). To appear.

[31] Laura Grigori, Sophie Moufawad, and Frederic Nataf, *Enlarged Krylov subspace conjugate gradient methods for reducing communication*, SIAM Jouranl on Matrix Analysis and Applications, 37 (2016), pp. 744–773.

[32] Wolter Groenevelt, Mourad E.H. Ismail, and Erik Koelink, *Spectral decomposition and matrix-valued orthogonal polynomials*, Advances in Mathematics, 244 (2013), pp. 91–105.

[33] Martin H. Gutknecht, *Block Krylov space methods for linear systems with multiple right-hand sides: an introduction*, in Modern Mathematical Models, Methods and Algorithms for Real World Systems, A.H. Siddiqi, I.S. Duff, and O. Christensen, eds., Anamaya, New Delhi, 2007, pp. 420–447.

[34] Martin H. Gutknecht and Thomas Schmelzer, *The block grade of a block Krylov space*, Linear Algebra and its Applications, 430 (2009), pp. 174–185.

[35] Peter Henrici, *Applied and Computational Complex Analysis, volume 2*, John Wiley & Sons, New York, 1977.

[36] Mohammed Heyouni and Aseddine Essai, *Matrix Krylov subspace methods for linear systems with multiple right-hand sides*, Numerical Algorithms, 40 (2005), pp. 137–156.

[37] Nicholas J. Higham, *Functions of Matrices: Theory and Computation*, SIAM, Philadelphia, 2008.

[38] Mark Hoemmen, *Communication-Avoiding Krylov Subspace Methods*, Ph.D. Thesis, University of California,

Berkeley, 2010.

[39] Milos Ilić, Ian W. Turner, and Daniel P. Simpson, *A restarted Lanczos approximation to functions of a symmetric matrix*, IMA Journal of Numerical Analysis, 30 (2010), pp. 1044–1061.

[40] Khalide Jbilou, Abderrahim Messaoudi, and Hassane Sadok, *Global FOM and GMRES algorithms for matrix equations*, Applied Numerical Mathematics, 31 (1999), pp. 49–63.

[41] Leonid Knizhnerman and Valeria Simoncini, *A new investigation of the extended Krylov subspace method for matrix function evaluations*, Numerical Linear Algebra and Applications, 17 (2010), pp. 615–638.

[42] Peter Lancaster, *Lambda-matrices and Vibrating Systems*, Pergamon Press, Oxford, 1966.

[43] Luciano Lopez and Valeria Simoncini, *Preserving geometric properties of the exponential matrix by block Krylov subspace methods*, BIT Numerical Mathematics, 46 (2006), pp. 813–830.

[44] David G. Luenberger, *Linear and Nonlinear Programming*, Adison-Wesley, Reading, Massachusetts, 2nd ed., 1984.

[45] Ronald B. Morgan, *Restarted block-GMRES with deflation of eigenvalues*, Applied Numerical Mathematics, 54 (2005), pp. 222–236.

[46] Herbert Neuberger, *Exactly massless quarks on the lattice*, Physics Letters B, 417 (1998), p. 141144.

[47] Andy A. Nikishin and Alex Yu. Yeremin, *Variable block CG algorithms for solving large sparse symmetric positive definite linear systems on parallel computers, I: General iterative scheme*, SIAM Journal of Matrix Analysis and Applications, 16 (1995), pp. 1135–1153.

[48] ———, *An automatic procedure for updating the block size in the block conjugate gradient method for solving linear systems*, Journal of Mathematical Sciences, 114 (2003), pp. 1844–1853.

[49] Dianne P. O'Leary, *The block conjugate gradient algorithm and related methods*, Linear Algebra and its Applications, 29 (1980), pp. 293–322.

[50] Michael L. Parks, Eric de Sturler, Greg Mackey, Duane D. Johnson, and Spandan Maiti, *Recycling Krylov subspaces for sequences of linear systems*, SIAM Journal of Scientific Computing, 28 (2006), pp. 1651–1674.

[51] Somaiyeh Rashedi, Ghodrat Ebadi, Sebastian Birk, and Andreas Frommer, *On short recurrence Krylov type methods for linear systems with many right-hand sides*, Journal of Computational and Applied Mathematics, 200 (2016), pp. 18–29.

[52] Youcef Saad, *On the Lanczos method for solving symmetric linear systems with several right-hand sides*, Mathematics of Computation, 47 (1987), pp. 651–662.

[53] Yousef Saad, *Analysis of some Krylov subspace approximations to the matrix exponential operator*, SIAM Journal on Numerical Analysis, 29 (1992), pp. 209–228.

[54] ———, *Iterative Methods for Sparse Linear Systems*, SIAM, Philadelphia, 2nd ed., 2003.

[55] Thomas Schmelzer, *Block Krlyov methods for Hermitian linear systems*, Ph.D. Thesis, Department of Mathematics, University of Kaiserslautern, 2004.

[56] Marcel Schweitzer, *Restarting and error estimation in polynomial and extended Krylov subspace methods for the approximation of matrix functions*, Ph.D. Thesis, Bergische Universität Wuppertal, 2016.

[57] Valeria Simoncini, *Ritz and pseudo-Ritz values using matrix polynomials*, Linear Algebra and its Applications, 241-243 (1996), pp. 787–801.

[58] Valeria Simoncini and Efstratios Gallopoulos, *An iterative method for nonsymmetric systems with multiple right-hand sides*, SIAM Journal on Scientific Computing, 16 (1995), pp. 917–933.

[59] ———, *Convergence properties of block GMRES and matrix polynomials*, Linear Algebra and its Applications, 247 (1996), pp. 97–119.

[60] ———, *A hybrid block GMRES method for nonsymmetric systems with multiple right-hand sides*, Journal of Computational and Applied Mathematics, 66 (1996), pp. 457–469.

[61] Valeria Simoncini and Daniel B. Szyld, *Recent computational developments in Krylov subspace methods for linear systems*, Numerical Linear Algebra with Applications, 14 (2007), pp. 1–59.

[62] Kirk M. Soodhalter, *A block MINRES algorithm based on the band Lanczos method*, Numerical Algorithms, 69 (2015), pp. 473–494.

[63] Lloyd N. Trefethen, J. A. C. Weideman, and Thomas Schmelzer, *Talbot quadratures and rational approximations*, BIT, 46 (2006), pp. 653–670.

[64] J. A. C. Weideman, *Optimizing Talbot's contours for the inversion of the Laplace transform*, SIAM J. Numer. Anal., 44 (2006), pp. 2342–2362.

[65] J. A. C. Weideman and Lloyd N. Trefethen, *Parabolic and hyperbolic contours for computing the Bromwich integral*, Math. Comp., 76 (2007), pp. 1341–1356.

[66] Jianhua Zhang, Hua Dai, and Jing Zhao, *A new family of global methods for linear systems with multiple right-hand sides*, Journal of Computational and Applied Mathematics, 236 (2011), pp. 1562–1575.