ITERATIVE TIKHONOV REGULARIZATION OF TENSOR EQUATIONS BASED ON THE ARNOLDI PROCESS AND SOME OF ITS GENERALIZATIONS

FATEMEH PANJEH ALI BEIK[1], MEHDI NAJAFI–KALYANI[1] AND LOTHAR REICHEL[2]

Abstract. We consider the solution of linear discrete ill-posed systems of equations with a certain tensor product structure. Two aspects of this kind of problems are investigated: They are transformed to large linear systems of equations and the conditioning of the matrix of the latter system is analyzed. Also, the distance of this matrix to symmetry and skew-symmetry is investigated. The aim of our analysis is to shed light on properties of linear discrete ill-posed problems and to study the feasibility of using Krylov subspace iterative methods in conjunction with Tikhonov regularization to solve Sylvester tensor equations with severely ill-conditioned coefficient matrices. The performance of several proposed algorithms is studied numerically. Applications include color image restoration and the solution of a 3D radiative transfer equation that is discretized by a Chebyshev collocation spectral method.

Key words. Generalized Arnoldi process; global Arnoldi process; flexible Arnoldi process; Sylvester tensor equation; ill-posed problem; Tikhonov regularization.

AMS subject classifications. 65F10; 15A69; 65F22.

1. Introduction.

This paper discusses the solution of severely ill-conditioned tensor equations that arise in color image restoration, video restoration, and when solving certain partial differential equations in several space-dimensions by collocation methods. A tensor is a multidimensional array. The number of indices of its entries is referred to as mode or way. Throughout this paper vectors (tensors of order one) and matrices (tensors of order two) are denoted by lower case and upper case letters, respectively; Euler script letters stand for tensors of order three or higher. The element $(i_1, i_2, \ldots, i_N)$ of an N-mode tensor $\mathcal{X}$ is denoted by $x_{i_1 i_2 \ldots i_N}$.

Consider the Sylvester tensor equation

$$\mathcal{X} \times_1 A^{(1)} + \mathcal{X} \times_2 A^{(2)} + \ldots + \mathcal{X} \times_N A^{(N)} = \mathcal{D}, \tag{1.1}$$

where the right-hand side tensor $\mathcal{D} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$ and the coefficient matrices $A^{(n)} \in \mathbb{R}^{I_n \times I_n}$ $(n = 1, 2, \ldots, N)$ are known, and $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$ is the unknown tensor to be determined. The definition of the n-mode product $\times_n$ is the standard one, see, e.g., [27]; details are given in Subsection 1.1. Equations of the form (1.1) arise from the discretization of a linear partial differential equation in several space-dimensions by finite differences [3, 4, 5, 7, 11] or by spectral methods [5, 29, 34, 35, 36, 45]. We refer the reader to [25] for a survey of tensor numerical methods for the solution of partial differential equations in many space-dimensions. Equations of the form (1.1) also arise in the restoration of color and hyperspectral images, see, e.g., [6, 17, 30, 43], blind source separation [31], and when describing a chain of spin particles [1].

Krylov subspace methods are popular solution methods for Sylvester tensor equations (1.1). For the case when the right-hand side in (1.1) is a tensor of low rank, Krylov subspace methods have been studied by Kressner and Tobler [28]. Ballani and Grasedyck [4] implemented the GMRES method with Hierarchical Tucker Format (HTF) tensor truncation and multigrid acceleration. Chen and Lu [11] proposed the GMRES method based on tensor format (GMRES_BTF) for solving (1.1) in the situation when the right-hand side is not necessarily a low-rank tensor. In [5], the tensor form of the FOM algorithm (FOM_BTF) was proposed. Also a nested algorithm for the situation when Eq. (1.1) has nonsymmetric

[1] Department of Mathematics, Vali-e-Asr University of Rafsanjan, PO Box 518, Rafsanjan, Iran (f.beik@vru.ac.ir (F. P. A. Beik); m.najafi.uk@gmail.com (Mehdi Najafi–Kalyani)).

[2] Department of Mathematical Sciences, Kent State University, Kent, OH 44242, USA (reichel@math.kent.edu).

positive definite coefficient matrices was examined. We also refer to Fan et al. [16] for a recent discussion on solution methods for Sylvester tensor equations (1.1) that arise from the discretization of elliptic partial differential equations in higher space-dimensions.

To the best of our knowledge, the performance of iterative methods tailored for the solution of problems of the form (1.1) has not received much attention in the literature so far. We are primarily concerned with the situation when the equation stems from the discretization of a linear ill-posed problem. Then (1.1) is referred to as a discrete ill-posed problem. Such problems arise, e.g., in color image restoration. In this application the right-hand side is typically contaminated by an error $\mathcal{E}$, i.e.,

$$\mathcal{D} = \tilde{\mathcal{D}} + \mathcal{E}, \tag{1.2}$$

where $\tilde{\mathcal{D}}$ denotes the unknown error-free right-hand side. It represents a blurred, but noise-free, image.

We would like to determine the solution, denoted by $\tilde{\mathcal{X}}$, of minimum norm (to be defined) of the tensor equation (1.1) with the right-hand side replaced by $\tilde{\mathcal{D}}$, i.e., of the unavailable Sylvester tensor equation

$$\mathcal{X} \times_1 A^{(1)} + \mathcal{X} \times_2 A^{(2)} + \ldots + \mathcal{X} \times_N A^{(N)} = \tilde{\mathcal{D}}. \tag{1.3}$$

This equation is assumed to be consistent, but equation (1.1) does not have to be. Since the right-hand side $\tilde{\mathcal{D}}$ is not known, we may try to determine an approximation of $\tilde{\mathcal{X}}$ by solving (1.1) with an available iterative method for the solution of Sylvester tensor equations, e.g., one of the methods described in [4, 5, 11, 16, 28]. However, when (1.1) is a discrete ill-posed problem, the computed solution so obtained is likely to be a poor approximation of $\tilde{\mathcal{X}}$ due to severe propagation of the error $\mathcal{E}$ in $\mathcal{D}$ into the computed solution. It is the aim of the present paper to generalize results and techniques in [6, 8, 38] to overcome this difficulty. This leads us to a Tikhonov regularization strategy, in which the problem of solving (1.1) is replaced by the solution of a minimization problem of the form

$$\min_{\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}} \left\{ \left\| \sum_{i=1}^{N} \mathcal{X} \times_i A^{(i)} - \mathcal{D} \right\|^2 + \lambda \left\| \sum_{j=1}^{M} \mathcal{X} \times_j L^{(j)} \right\|^2 \right\}, \tag{1.4}$$

where $1 \leq M \leq N$ and the $L^{(j)}$ $(j = 1, 2, \ldots, M)$ are regularization matrices. The nonnegative constant λ is a regularization parameter. Note that N stands for the number of modes in the unknown tensor $\mathcal{X}$. Throughout this paper, we will measure a tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$ with the norm

$$\|\mathcal{X}\| := \sqrt{\sum_{i_1=1}^{I_1} \sum_{i_2=1}^{I_2} \cdots \sum_{i_N=1}^{I_N} x_{i_1 i_2 \ldots i_N}^2},$$

which generalizes the matrix Frobenius norm.

It is well-known that (1.1) is equivalent to the following linear system of equations

$$\mathcal{A} x = b, \tag{1.5}$$

with $x = \mathrm{vec}(\mathcal{X})$, $b = \mathrm{vec}(\mathcal{D})$, and

$$\mathcal{A} = \sum_{j=1}^{N} I^{(I_N)} \otimes \ldots \otimes I^{(I_{j+1})} \otimes A^{(j)} \otimes I^{(I_{j-1})} \otimes \ldots \otimes I^{(I_1)}. \tag{1.6}$$

Here and throughout this paper, "vec" stands for the standard vectorization operator that transforms a tensor to a vector. We note that $\mathrm{vec}(\mathcal{X})$ is obtained by using the standard vectorization operator with respect to frontal slices of $\mathcal{X}$. Recall that for a given tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$, the frontal slices are defined by

$$\mathcal{X}_{\underbrace{::\ldots:}_{(N-1)\text{-times}}k} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_{N-1}}, \quad k = 1, 2, \ldots, I_N,$$

which also are known as column tensors of $\mathcal{X}$. The kth frontal slice of $\mathcal{X}$ is obtained by setting the last index to k.

We primarily consider the situation when the solution of (1.1) (or equivalently of (1.5)) is an ill-posed problem. Therefore, we first investigate the dependence of the condition number of $\mathcal{A}$ on the condition numbers of the matrices $A^{(j)}$ in (1.6). We are interested in the behavior of iterative methods applied to the solution of (1.1) (or equivalently to the solution of (1.5)). The behavior of iterative methods for the solution of large linear systems of equations (1.5) with a symmetric matrix $\mathcal{A}$ is better understood than the behavior of iterative methods applied to the solution of linear systems of equations with a nonsymmetric matrix. When solving (1.5) by the GMRES iterative method, which is based on the Arnoldi process (to be defined below), the distance of $\mathcal{A}$ to the set of symmetric and to the set of skew-symmetric matrices is important; see [18]. The methods we consider for the solution of (1.1) are based on the Arnoldi process. We are therefore interested in how the distance of the matrices $A^{(i)}$, $i = 1, 2, \ldots, N$, to the sets of symmetric and skew-symmetric matrices affects the behavior of the iterative methods considered. We will study this by introducing (fairly) easily computable distance measures for the matrix (1.6). A generalization of (1.6) that arises in color image restoration with cross-channel blur also will be discussed.

This paper is organized as follows. In the remainder of this section, we review some basic concepts and introduce notation used in later sections. In Section 2, which is motivated by results in [33, 42], we derive lower and upper bounds for the condition number of $\mathcal{A}$, given by (1.6), in terms of extreme singular values of the matrices $A^{(i)}$ for $i = 1, 2, \ldots, N$. Section 3 is concerned with measuring the distance of a matrix with Kronecker structure, that is associated with (1.1), to the set of symmetric (positive or negative semi-definite) matrices, or to the set of skew-symmetric matrices. A new distance measure is introduced that allows efficient computation. The aim of Section 4 is to present iterative methods based on Arnoldi-type processes that exploit the tensor structure to solve (1.4). To this end, we apply results in [5] and extend techniques that have been described in [6, 8, 24, 38]. Numerical results that illustrate the results of Sections 2 and 3, as well as the effectiveness of the proposed iterative schemes are reported in Section 5. Concluding remarks can be found in Section 6.

1.1. Preliminaries.

This subsection briefly reviews some basic definitions and properties that are used in the remainder of the paper. Our notation follows [27].

The inner product between two tensors of the same size $\mathcal{X}, \mathcal{Y} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$ is defined by

$$\langle \mathcal{X}, \mathcal{Y} \rangle := \sum_{i_1=1}^{I_1} \sum_{i_2=1}^{I_2} \cdots \sum_{i_N=1}^{I_N} x_{i_1 i_2 \ldots i_N} y_{i_1 i_2 \ldots i_N}.$$

The n-mode (matrix) product of a tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$ with a matrix $U \in \mathbb{R}^{J \times I_n}$ is denoted by $\mathcal{X} \times_n U$. It is of size $I_1 \times \ldots \times I_{n-1} \times J \times I_{n+1} \times \ldots \times I_N$ and its elements are given by

$$(\mathcal{X} \times_n U)_{i_1 \cdots i_{n-1} j i_{n+1} \cdots i_N} = \sum_{i_n=1}^{I_n} x_{i_1 i_2 \ldots i_N} u_{j i_n}, \quad j = 1, 2, \ldots, J.$$

The n-mode (vector) product of a tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$ with a vector $v \in \mathbb{R}^{I_n}$ is of order $N-1$ and is denoted by $\mathcal{X} \bar{\times}_n v$, where its size is given by $I_1 \times \ldots \times I_{n-1} \times I_{n+1} \times \ldots \times I_N$.

We will use the $\boxtimes^N$ product between two N-mode tensors $\mathcal{X}$ and $\mathcal{Y}$, which is a reformulation of a special case of the contracted product. In fact, the product $\mathcal{X} \boxtimes^N \mathcal{Y}$ is the contracted product of N-mode tensors $\mathcal{X}$ and $\mathcal{Y}$ along the first $N-1$ modes; see [5, 12] for further details.

We conclude this subsection with a proposition that can be established by using the definitions of n-mode and contracted products. This result is useful for deriving iterative methods in the tensor framework.

PROPOSITION 1.1. *Suppose that* $\mathcal{B} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N \times m}$ *is an* $(N+1)$*-mode tensor with column tensors* $\mathcal{B}_1, \mathcal{B}_2, \ldots, \mathcal{B}_m \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$ *and let* $z = (z_1, z_2, \ldots, z_m)^T \in \mathbb{R}^m$*. For an arbitrary* $(N+1)$*-mode tensor* $\mathcal{A}$ *with* N*-mode column tensors* $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_m$*, we have*

$$\mathcal{A} \boxtimes^{(N+1)} (\mathcal{B} \bar{\times}_{N+1} z) = (\mathcal{A} \boxtimes^{(N+1)} \mathcal{B}) z,$$
$$(\mathcal{A} \bar{\times}_{N+1} z) \boxtimes^{(N+1)} \mathcal{B} = z^T (\mathcal{A} \boxtimes^{(N+1)} \mathcal{B}).$$

1.2. Notation.

For a real square matrix A with real eigenvalues, $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ denote the minimum and maximum eigenvalues of A, respectively. Further, $\lambda(A)$ stands for an arbitrary eigenvalue of A, and the set of all eigenvalues of A is denoted by $\sigma(A)$. The identity matrix of order n is denoted by $I^{(n)}$, and the vector e_i stands for the ith column of an identity matrix of suitable order.

The symmetric and skew-symmetric parts of a real square matrix A are given by

$$\mathcal{H}(A) := \frac{1}{2}(A + A^T) \quad \text{and} \quad \mathcal{S}(A) := \frac{1}{2}(A - A^T), \tag{1.7}$$

respectively, where the superscript ^{T} denotes transposition. The condition number of a (square) invertible matrix A is defined as $\mathrm{cond}(A) := \|A\|_2 \|A^{-1}\|_2$, where $\|\cdot\|_2$ denotes the matrix spectral norm. We write $A \succ 0$ ($A \succcurlyeq 0$) to indicate that the matrix A is symmetric positive (semi-)definite.

Let B be an arbitrary matrix. The maximum and minimum singular values of B are denoted by $\sigma_{\max}(B)$ and $\sigma_{\min}(B)$, respectively. The notation $\mathrm{Null}(B)$ stands for the null space of B.

The Kronecker product of two matrices $X = [x_{ij}] \in \mathbb{R}^{n \times p}$ and $Y \in \mathbb{R}^{q \times l}$ is defined by $X \otimes Y = [x_{ij} Y] \in \mathbb{R}^{nq \times pl}$.

2. Conditioning.

The Kronecker structure of the matrix $\mathcal{A}$ defined by (1.6) makes it difficult to analyze the problem (1.5). In particular, it is difficult to approximate the inverse of $\mathcal{A}$ for general matrices $A^{(i)}$ already for $N = 2$. This has recently been pointed out by Simoncini [44, Section 9]. Nevertheless, some insight can be gained by investigating the condition number of $\mathcal{A}$. This section derives lower and upper bounds for $\mathrm{cond}(\mathcal{A})$. The bounds obtained are helpful for discussing the conditioning of (1.1).

Let $\tilde{\mathcal{X}}$ solve the Sylvester tensor equation with error-free right-hand side (1.3). Shi et al. [42, p. 1443] obtained the relative error bound

$$\frac{\|\tilde{\mathcal{X}} - \mathcal{X}\|}{\|\tilde{\mathcal{X}}\|} \leq \sum_{i=1}^{N} \|A^{(i)}\|_F \frac{\prod_{i=1}^{N} \mathrm{cond}(T_i)}{\min\limits_{\lambda_i \in \sigma(A^{(i)})} |\sum_{i=1}^{N} \lambda_i|} \frac{\|\tilde{\mathcal{D}} - \mathcal{D}\|}{\|\tilde{\mathcal{D}}\|} \tag{2.1}$$

for the case when the matrices $A^{(i)}$ are diagonalizable, i.e., when there are nonsingular matrices T_i and diagonal matrices Λ_i such that $T_i^{-1} A^{(i)} T_i = \Lambda_i$ for $i = 1, 2, \ldots, N$. The norm $\|\cdot\|_F$ denotes the Frobenius matrix norm.

Let the matrix A be positive stable, i.e., all of its eigenvalues lie in the open right half plane. Liang and Zheng [33, p. 8] considered the case when all the matrices $A^{(i)}$ are equal to the same positively stable matrix A. They established the relative error bound

$$\frac{\|\tilde{\mathcal{X}}-\mathcal{X}\|}{\|\tilde{\mathcal{X}}\|}\leq N\|A\|_{_F}\|\mathcal{A}^{-1}\|_{_2}\frac{\|\tilde{\mathcal{D}}-\mathcal{D}\|}{\|\tilde{\mathcal{D}}\|}. \tag{2.2}$$

The bounds (2.1) and (2.2) are valid when there are no perturbations in the coefficient matrices $A^{(i)}$. The identities

$$\frac{\|\tilde{\mathcal{X}}-\mathcal{X}\|}{\|\tilde{\mathcal{X}}\|}=\frac{\|\mathrm{vec}(\tilde{\mathcal{X}})-\mathrm{vec}(\mathcal{X})\|_{_2}}{\|\mathrm{vec}(\tilde{\mathcal{X}})\|_{_2}}\quad\text{and}\quad\frac{\|\tilde{\mathcal{D}}-\mathcal{D}\|}{\|\tilde{\mathcal{D}}\|}=\frac{\|\mathrm{vec}(\tilde{\mathcal{D}})-\mathrm{vec}(\mathcal{D})\|_{_2}}{\|\mathrm{vec}(\tilde{\mathcal{D}})\|_{_2}}$$

show that perturbation analysis for (1.1) is closely related to obtaining bounds for the condition number of $\mathcal{A}$.

We would like to derive lower and upper bounds for cond($\mathcal{A}$) in terms of singular values of the matrices $A^{(i)}$ under some sufficient conditions. These kinds of bounds can be cheaply computed since the sizes of the matrices $A^{(i)}$ are small in comparison to the size of $\mathcal{A}$. The main challenge is to determine bounds for $\|\mathcal{A}^{-1}\|_{_2}$. Let us first recall the following result, which is an immediate consequence of Weyl's Theorem; see [23, Theorem 4.3.1].

Proposition 2.1. *Let $A,B\in\mathbb{R}^{n\times n}$ be symmetric matrices. Then*

$$\lambda_{\max}(A+B)\leq\lambda_{\max}(A)+\lambda_{\max}(B)\quad\textit{and}\quad\lambda_{\min}(A+B)\geq\lambda_{\min}(A)+\lambda_{\min}(B).$$

We derive a lower bound for the condition number of $\mathcal{A}$ by applying bounds for the extreme eigenvalues of $\mathcal{A}\mathcal{A}^T$. The following result will be used to achieve this goal.

Proposition 2.2. *Let $A\in\mathbb{R}^{n\times n}$ and $B\in\mathbb{R}^{m\times m}$. Then*

$$(x^T\otimes y^T)\mathcal{H}(A\otimes B)(x\otimes y)=(x^T\mathcal{H}(A)x)\times(y^T\mathcal{H}(B)y)$$

for any vectors $x\in\mathbb{R}^n$ and $y\in\mathbb{R}^m$.

Proof. From [46], it is known that $\mathcal{H}(A\otimes B)=\mathcal{H}(A)\otimes\mathcal{H}(B)+\mathcal{S}(A)\otimes\mathcal{S}(B)$. The proof now follows from the fact that $x^T\mathcal{S}(A)x=0$ and $y^T\mathcal{S}(B)y=0$. □

Proposition 2.3. *Assume that $\mathcal{A}$ and $A^{(i)}$ are invertible matrices for $i=1,2,\ldots,N$. Then*

$$\frac{1}{\|\mathcal{A}^{-1}\|_2^2}=\lambda_{\min}(\mathcal{A}\mathcal{A}^T)\leq\left(\sum\nolimits_{i=1}^N\sigma_{\min}(A^{(i)})\right)^2 \tag{2.3}$$

and

$$\lambda_{\max}(\mathcal{A}\mathcal{A}^T)\geq\sum_{i=1}^N\sigma_{\max}^2(A^{(i)})+2\sum_{i=1}^N\sum_{j=i+1}^N(y_i\otimes y_j)^T\mathcal{H}(A^{(i)}\otimes A^{(j)})(y_i\otimes y_j), \tag{2.4}$$

where $A^{(i)}(A^{(i)})^Ty_i=\sigma_{\max}^2(A^{(i)})y_i$ with $\|y_i\|_2=1$ for $i=1,2,\ldots,N$.

Proof. For simplicity, we show the validity of the assertion for $N=3$. A similar strategy can be used to show (2.3) and (2.4) for an arbitrary integer $N\geq1$. For notational convenience, set $A^{(1)}=A$, $A^{(2)}=B$, and $A^{(3)}=C$. Let $\sigma_{\min}^2(A)$, $\sigma_{\min}^2(B)$, and $\sigma_{\min}^2(C)$ stand for the minimal eigenvalues of AA^T, BB^T, and CC^T, respectively, and let x, y, and z denote corresponding unit eigenvectors.

Let $\mathcal{Y}=z\otimes y\otimes x$. Straightforward computations and the Cauchy–Schwartz inequality give

$$\begin{aligned}(z^T\otimes y^T\otimes x^T)(I\otimes B^T\otimes A)(z\otimes y\otimes x)&=\|z\|_2^2\times\langle By,y\rangle\times\langle x,Ax\rangle\\&=\sigma_{\min}^2(B)\langle y,B^{-1}y\rangle\times\sigma_{\min}^2(A)\langle A^{-1}x,x\rangle\\&\leq\sigma_{\min}(B)\sigma_{\min}(A).\end{aligned} \tag{2.5}$$

By computing the quadratic form associated with $\mathcal{A}\mathcal{A}^T$, it can be seen that

$$\begin{aligned}\mathcal{Y}^T\mathcal{A}\mathcal{A}^T\mathcal{Y} = &\langle AA^Tx,x\rangle\|y\|_2^2\|z\|_2^2 + \langle BB^Ty,y\rangle\|x\|_2^2\|z\|_2^2 + \langle CC^Tz,z\rangle\|x\|_2^2\|y\|_2^2\\ &+ \langle By,y\rangle\times\langle x,Ax\rangle\|z\|_2^2 + \langle y,By\rangle\times\langle Ax,x\rangle\|z\|_2^2 + \langle z,Cz\rangle\times\langle By,y\rangle\|x\|_2^2\\ &+ \langle Cz,z\rangle\times\langle y,By\rangle\|x\|_2^2 + \langle Cz,z\rangle\times\langle x,Ax\rangle\|y\|_2^2 + \langle z,Cz\rangle\times\langle Ax,x\rangle\|y\|_2^2.\end{aligned}$$

Since x,y,z are unit eigenvectors for AA^T, BB^T and CC^T, respectively, with similar computations used for deriving (2.5), we obtain

$$\begin{aligned}\mathcal{Y}^T\mathcal{A}\mathcal{A}^T\mathcal{Y} \leq\ &\sigma_{\min}^2(A) + \sigma_{\min}^2(B) + \sigma_{\min}^2(C)\\ &+2\sigma_{\min}(A)\sigma_{\min}(B) + 2\sigma_{\min}(A)\sigma_{\min}(C) + 2\sigma_{\min}(B)\sigma_{\min}(C)\\ =\ &\left(\sigma_{\min}(A) + \sigma_{\min}(B) + \sigma_{\min}(C)\right)^2.\end{aligned}$$

As a result, using the fact that $\lambda_{\min}(\mathcal{A}\mathcal{A}^T) \leq \mathcal{Y}^T\mathcal{A}\mathcal{A}^T\mathcal{Y}$, the validity of (2.3) can be deduced.

We turn to the proof of (2.4). Let $\sigma_{\max}^2(A)$, $\sigma_{\max}^2(B)$, and $\sigma_{\max}^2(C)$ denote the maximum eigenvalues of the matrices AA^T, BB^T, and CC^T, respectively, and let $\tilde{x}$, $\tilde{y}$, and $\tilde{z}$ stand for associated unit eigenvectors. Setting $\tilde{\mathcal{Y}} = \tilde{z}\otimes\tilde{y}\otimes\tilde{x}$, it can be seen that

$$\begin{aligned}\tilde{\mathcal{Y}}^T\mathcal{A}\mathcal{A}^T\tilde{\mathcal{Y}} =&\sigma_{\max}^2(A) + \sigma_{\max}^2(B) + \sigma_{\max}^2(C) + 2\tilde{y}^T\mathcal{H}(B)\tilde{y}\times\tilde{x}^T\mathcal{H}(A)\tilde{x}\\ &+ 2\tilde{z}^T\mathcal{H}(C)\tilde{z}\times\tilde{x}^T\mathcal{H}(A)\tilde{x} + 2\tilde{z}^T\mathcal{H}(C)\tilde{z}\times\tilde{y}^T\mathcal{H}(B)\tilde{y}.\end{aligned}$$

Using Proposition 2.2 and the fact that $\lambda_{\max}(\mathcal{A}\mathcal{A}^T) \geq \tilde{\mathcal{Y}}^T\mathcal{A}\mathcal{A}^T\tilde{\mathcal{Y}}$, it is not difficult to verify (2.4).
□

Remark 2.4. From the proof of the previous proposition, we may immediately conclude that

$$\mathrm{cond}(\mathcal{A}) \geq \frac{\sqrt{\sum_{i=1}^N \sigma_{\max}^2(A^{(i)}) + 2\sum_{i=1}^N\sum_{j=i+1}^N \left(y_i^T\mathcal{H}(A^{(i)})y_i\right)\left(y_j^T\mathcal{H}(A^{(j)})y_j\right)}}{\sum\limits_{i=1}^N \sigma_{\min}(A^{(i)})},$$

where $A^{(i)}(A^{(i)})^Ty_i = \sigma_{\max}^2(A^{(i)})y_i$ with $\|y_i\|_2 = 1$ for $i = 1,2,\ldots,N$. By adding some mild conditions to the assumptions of Proposition 2.3, we obtain simpler lower bounds for $\mathrm{cond}(\mathcal{A})$ in terms of the singular values of the matrices $A^{(1)}, A^{(2)}, \ldots, A^{(N)}$. For instance,

- if all the matrices are equal, i.e., if $A^{(i)} = A$ for $i = 1,2,\ldots,N$, then

$$\mathrm{cond}(\mathcal{A}) \geq \frac{\sigma_{\max}(A)}{\sqrt{N}\sigma_{\min}(A)};$$

- if either of the following statements is true:
  a) the (possibly nonsymmetric) matrices $A^{(1)}, A^{(2)}, \ldots, A^{(N)}$ are all positive definite, or
  b) the matrices of the form $A^{(i)}\otimes A^{(j)}$ for $i,j = 1,2,\ldots,N$ and $i \neq j$ are positive definite,
  then

$$\mathrm{cond}(\mathcal{A}) \geq \frac{\sqrt{\sum\limits_{i=1}^N \sigma_{\max}^2(A^{(i)})}}{\sum\limits_{i=1}^N \sigma_{\min}(A^{(i)})}.$$

To derive an upper bound for the condition number of $\mathcal{A}$, we require additional conditions on the coefficient matrices $A^{(i)}$. Specifically, we need a condition that ensures the positive definiteness of the matrices $A^{(i)} \otimes (A^{(j)})^T$ for $i, j = 1, 2, \ldots, N$. To this end, we recall the following remark from [46].

Remark 2.5. Let F and G be two nonsymmetric matrices. If

$$\lambda_{\min}(\mathcal{H}(G))\lambda_{\min}(\mathcal{H}(F)) + \min\left(-\lambda(\mathcal{S}(G))\lambda(\mathcal{S}(F))\right) > 0,$$

then the symmetric part of $G^T \otimes F$ is positive definite.

We now establish bounds for the extreme eigenvalues of $\mathcal{A}\mathcal{A}^T$. These bounds will be used to obtain an upper bound for $\text{cond}(\mathcal{A})$. For notational simplicity, we assume that $N = 3$ and let $A = A^{(1)}$, $B = A^{(2)}$, and $C = A^{(3)}$. Analogues of the bounds (2.6) and (2.7) can be shown in a similar fashion when N is a general positive integer.

Proposition 2.6. *Let $\mathcal{A} = (I \otimes I \otimes A) + (I \otimes B \otimes I) + (C \otimes I \otimes I)$. Then*

$$\lambda_{\max}(\mathcal{A}\mathcal{A}^T) \leq (\sigma_{\max}(A) + \sigma_{\max}(B) + \sigma_{\max}(C))^2. \tag{2.6}$$

Moreover, assume that $\mathcal{A}$ is invertible. Then, if $B^T \otimes A$, $C^T \otimes A$, and $C^T \otimes B$ are positive definite, we have

$$\frac{1}{\lambda_{\min}(\mathcal{A}\mathcal{A}^T)} \leq \frac{1}{\sigma^2_{\min}(A) + \sigma^2_{\min}(B) + \sigma^2_{\min}(C)}. \tag{2.7}$$

Proof. The spectral norm of $\mathcal{A} = (I \otimes I \otimes A) + (I \otimes B \otimes I) + (C \otimes I \otimes I)$ and the triangle inequality give (2.6). To show (2.7), we first note that $C \otimes I \otimes A^T$ is congruent to $I \otimes A^T \otimes C$. Hence, the positive definiteness of $B^T \otimes A$, $C^T \otimes A$, and $C^T \otimes B$ implies that the symmetric parts of the matrices $I \otimes B^T \otimes A$, $C^T \otimes B \otimes I$, and $C^T \otimes I \otimes A$ are positive definite. Some computations and Proposition 2.1 yield

$$\begin{aligned}\lambda_{\min}(\mathcal{A}\mathcal{A}^T) &\geq \sigma^2_{\min}(A) + \sigma^2_{\min}(B) + \sigma^2_{\min}(C) + \lambda_{\min}\left((I \otimes B^T \otimes A) + (I \otimes B \otimes A^T)\right) \\ &\quad + \lambda_{\min}\left((C^T \otimes B \otimes I) + (C \otimes B^T \otimes I)\right) + \lambda_{\min}\left((C^T \otimes I \otimes A) + (C \otimes I \otimes A^T)\right) \\ &\geq \sigma^2_{\min}(A) + \sigma^2_{\min}(B) + \sigma^2_{\min}(C).\end{aligned}$$

This completes the proof. □

Remark 2.7. The above proposition shows that

- if the matrices $B^T \otimes A$, $C^T \otimes A$, and $C^T \otimes B$ are positive definite, then we can determine an upper bound for $\text{cond}(\mathcal{A})$:

$$\text{cond}(\mathcal{A}) \leq \frac{\sigma_{\max}(A) + \sigma_{\max}(B) + \sigma_{\max}(C)}{\sqrt{\sigma^2_{\min}(A) + \sigma^2_{\min}(B) + \sigma^2_{\min}(C)}};$$

- if all the matrices $A^{(i)}$ that define $\mathcal{A}$ are all equal, i.e., if $A^{(1)} = \ldots = A^{(N)} = A$, and if $A \otimes A^T$ is positive definite, then

$$\text{cond}(\mathcal{A}) \leq \frac{\sqrt{N}\sigma_{\max}(A)}{\sigma_{\min}(A)}.$$

The following example reports some numerical experiments that illustrate the bounds of Remarks 2.4 and 2.7.

Example 2.8. *Let $N = 3$ in* (1.1) *and consider the two cases:*

Case 1: *All the coefficient matrices are equal to the $n \times n$ matrix*

$$A^{(1)} = A^{(2)} = A^{(3)} = M + 2rL + \frac{1}{(n+1)^2} I_n,$$

where $M = \text{tridiag}(-1, 2, -1)$, $L = \text{tridiag}(0.5, 0, -0.5)$, and $r = 0.01$. These matrices are discussed in [46].

TABLE 1
Results for Cases 1 and 2 of Example 2.8.

	Case 1			Case 2		
n	Lower bound	cond($\mathcal{A}$)	Upper bound	Lower bound	cond($\mathcal{A}$)	Upper bound
10	$2.54 \cdot 10^1$	$7.12 \cdot 10^1$	$7.61 \cdot 10^1$	$2.91 \cdot 10^1$	$8.46 \cdot 10^1$	$8.72 \cdot 10^1$
20	$9.32 \cdot 10^1$	$2.65 \cdot 10^2$	$2.79 \cdot 10^2$	$1.05 \cdot 10^2$	$3.05 \cdot 10^2$	$3.16 \cdot 10^2$
30	$2.03 \cdot 10^2$	$5.57 \cdot 10^2$	$6.08 \cdot 10^2$	$2.28 \cdot 10^2$	$6.60 \cdot 10^2$	$6.84 \cdot 10^2$
40	$3.54 \cdot 10^2$	$1.00 \cdot 10^3$	$1.06 \cdot 10^3$	$3.97 \cdot 10^2$	$1.14 \cdot 10^3$	$1.19 \cdot 10^3$
50	$5.55 \cdot 10^2$	$1.55 \cdot 10^3$	$1.63 \cdot 10^3$		–	
60	$7.75 \cdot 10^2$	$2.22 \cdot 10^3$	$2.32 \cdot 10^3$		–	
70	$1.04 \cdot 10^3$	$3.00 \cdot 10^3$	$3.12 \cdot 10^3$		–	

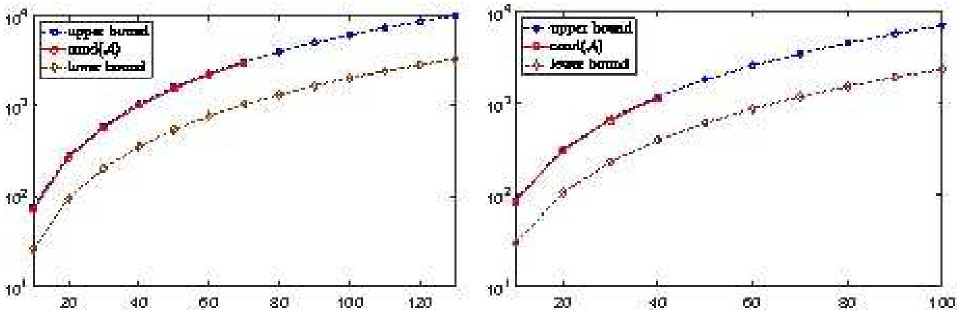

FIG. 1. *Lower and upper bounds for cond($\mathcal{A}$) versus the exact value: Case 1 (left) and Case 2 (Right).*

Case 2: *Regard the Sylvester tensor equation that arises from the discretization of a 3D convection-diffusion partial differential equation by standard finite differences on a uniform grid for the diffusion term and a second-order convergent scheme (Fromm's scheme) for the convection term. The coefficient matrices $A^{(i)}$ in* (1.1) *are given in [5, Example 5.4].*

Table 1 shows the bounds of Remarks 2.4 and 2.7. The symbol "–" in the table indicates that the computer used for the calculations[1] *was not capable of computing the condition number of $\mathcal{A}$ due to lack of memory. We remark that cond($\mathcal{A}$) is computed by using the MATLAB function* condest($\mathcal{A}$)*. Moreover, the condition number of $\mathcal{A}$ could not be computed for $n \geq 80$ in reasonable time in Case 1. Figure 1 depicts the computed upper and lower bounds together with the computed values of cond($\mathcal{A}$). The figure shows the lower and upper bounds for $n \leq 130$, whereas the values of cond($\mathcal{A}$) are only reported up to the largest value of n for which the condition number of $\mathcal{A} \in \mathbb{R}^{n^3 \times n^3}$ could be evaluated on our computer.*

Remark 2.9. The restoration of color images without cross-channel blur requires the solution of linear systems of equations (1.5) with a matrix of the form (1.6) with $N = 3$. The matrices $A^{(j)}$, $j = 1, 2, 3$, model within channel blur of channels that represent red, blue, and green light. Further details on color image restoration can be found, e.g., in [6]. In the presence of cross-channel blur, the matrix in the linear system of equations (1.5) is given by $\widetilde{\mathcal{A}} = C \otimes \mathcal{A}$, where

$$\mathcal{A} = I^{(3)} \otimes I^{(2)} \otimes A^{(1)} + I^{(3)} \otimes A^{(2)} \otimes I^{(1)} + A^{(3)} \otimes I^{(2)} \otimes I^{(1)}.$$

A bound for the condition number of the matrix $\widetilde{\mathcal{A}}$ can be obtained by applying the technique of this section to determine a bound for the condition number of $\mathcal{A}$ and using the fact that

$$\text{cond}(\widetilde{\mathcal{A}}) = \text{cond}(C)\text{cond}(\mathcal{A}).$$

[1] See Section 5 for details about the computer system.

The condition number of the small matrix C can be easily computed. Typically, matrices modeling cross-channel blur are not very ill-conditioned; see [6, Example 3] for an illustration.

3. Distance to symmetric semi-definiteness. For an arbitrary square matrix $A \in \mathbb{R}^{l\times l}$, the distances from A to the set of all symmetric positive semi-definite matrices and to the set of all negative semi-definite matrices are defined by

$$\delta^{+}(A) := \min\left\{\|E\| : \ E \in \mathbb{R}^{l\times l}, \ A+E \succcurlyeq 0\right\}$$

and

$$\delta^{-}(A) := \min\left\{\|E\| : \ E \in \mathbb{R}^{l\times l}, \ A+E \preccurlyeq 0\right\},$$

respectively, where $\|\cdot\|$ denotes a given matrix norm. Holmes (1974) and Higham (1988) derived expressions for the spectral and Frobenius norms, respectively. These expressions are provided in the following two theorems.

THEOREM 3.1. *[22] Let $A \in \mathbb{R}^{l\times l}$. Then*

$$\begin{aligned} \delta_{2}^{+}(A) &= \min\left\{\|E\|_{2} : \ E \in \mathbb{R}^{l\times l}, \ A+E \succcurlyeq 0\right\} \\ &= \min\left\{r \geq 0 : \ r^{2}I + (\mathcal{S}(A))^{2} \succcurlyeq 0 \ \textit{and}\ G(r) \succcurlyeq 0\right\}, \end{aligned} \tag{3.1}$$

where $G(r) := \mathcal{H}(A) + (r^{2}I + (\mathcal{S}(A)^{2})^{1/2}$. The matrix $P = G(\delta_{2}^{+}(A))$ is a positive semi-definite approximation of A in the spectral norm.

THEOREM 3.2. *[21, Theorem 2.1] Let $A \in \mathbb{R}^{l\times l}$ and let $\mathcal{H}(A) = UH$ be a polar decomposition. Then $X_F = (\mathcal{H}(A) + H)/2$ is the unique best positive approximation of A in the Frobenius norm. Moreover,*

$$\delta_{F}^{+}(A)^{2} = \sum_{\lambda_i(\mathcal{H}(A))<0} \left(\lambda_i\left(\mathcal{H}(A)\right)\right)^{2} + \|\mathcal{S}(A)\|_{F}^{2}. \tag{3.2}$$

Remark 3.3. Following a similar approach to the one used in [21, Theorem 2.1], one can show that

$$\delta_{F}^{-}(A)^{2} = \sum_{\lambda_i(\mathcal{H}(A))>0} \left(\lambda_i\left(\mathcal{H}(A)\right)\right)^{2} + \|\mathcal{S}(A)\|_{F}^{2}. \tag{3.3}$$

The aim of this section is to derive expressions for $\delta^{+}(\mathcal{A})$ and $\delta^{-}(\mathcal{A})$ for a suitable norm in the case when $\mathcal{A}$ is given in the Kronecker form (1.6). When determining $\delta_{F}^{\pm}(\mathcal{A})$ by using the coefficient matrices $A^{(i)}$ $(i = 1, 2, \ldots, N)$ and applying (3.1) and (3.2) (or (3.3)), this tends to be computationally expensive. We therefore propose to use an alternative norm.

Consider the real-valued function $\|\cdot\|_{ss}$ over the set of square matrices,

$$\|A\|_{ss} = \|\mathcal{H}(A)\|_{2} + \|\mathcal{S}(A)\|_{2},$$

where $\mathcal{H}(A)$ and $\mathcal{S}(A)$ are defined by (1.7). It is immediate to see that $\|\cdot\|_{ss}$ is a norm on the set of square matrices.

We will measure the distance of $A \in \mathbb{R}^{l\times l}$ to the set of positive and negative semi-definite matrices by

$$\delta_{ss}^{+}(A) = \min\left\{\|E\|_{ss} : \ E \in \mathbb{R}^{l\times l}, \ A+E \succcurlyeq 0\right\}$$

and

$$\delta_{ss}^{-}(A) = \min\left\{\|E\|_{ss} : \ E \in \mathbb{R}^{l\times l}, \ A+E \preccurlyeq 0\right\},$$

respectively, because of the relative ease of the computation of these quantities; see below.

We next derive expressions for $\delta^+_{ss}(A)$ and $\delta^-_{ss}(A)$ in terms of extreme eigenvalues of A. To this end, we first present the following result.

PROPOSITION 3.4. *Let the matrix $B \in \mathbb{R}^{l\times l}$ be symmetric. Then*

$$B + \|B\|_2 I \succcurlyeq 0 \quad \text{and} \quad B - \|B\|_2 I \preccurlyeq 0.$$

Proof. Let $\lambda \in \sigma(B)$. Then $|\lambda| \le \|B\|_2$ and $\lambda \pm \|B\|_2$ are eigenvalues of $B \pm \|B\|_2 I$. Therefore, the eigenvalues of $B + \|B\|_2 I$ are larger than or equal to zero, and the eigenvalues of $B - \|B\|_2 I$ are smaller than or equal to zero. This shows the assertion. □

The proof of the following proposition can be used to provide a simpler proof of [22, Theorem 1].

PROPOSITION 3.5. *For any square matrix A, the following statements hold:*

$$\delta^+_{ss}(A) = \max\{0, -\lambda_{\min}(\mathcal{H}(A))\} + \|\mathcal{S}(A)\|_2 \quad \text{and} \quad \delta^-_{ss}(A) = \max\{0, \lambda_{\max}(\mathcal{H}(A))\} + \|\mathcal{S}(A)\|_2 .$$

Proof. We only show the validity of the first equality; the expression for $\delta^-_{ss}(A)$ can be shown similarly. For any $X \succcurlyeq 0$, it follows from the symmetry of X that

$$\|A - X\|_{ss} = \|\mathcal{H}(A) - X\|_2 + \|\mathcal{S}(A)\|_2 ,$$

which shows that

$$\min_{X \succcurlyeq 0} \|A - X\|_{ss} = \min_{X \succcurlyeq 0} \|\mathcal{H}(A) - X\|_2 + \|\mathcal{S}(A)\|_2 .$$

To complete the proof, it suffices to show that

$$\min_{X \succcurlyeq 0} \|\mathcal{H}(A) - X\|_2 = \min\{r \ge 0 : \ \mathcal{H}(A) + rI \succcurlyeq 0\} . \tag{3.4}$$

Let $\tilde{r} = \|\mathcal{H}(A) - X\|_2$, where $X \succcurlyeq 0$ is given. Then $\mathcal{H}(A) - X + \tilde{r}I \succcurlyeq 0$ by Proposition 3.4. Therefore,

$$\tilde{r} \ge \min\{r \ge 0 : \ \mathcal{H}(A) + rI \succcurlyeq 0\} .$$

Since the above inequality holds for any $X \succcurlyeq 0$, we conclude that

$$\min_{X \succcurlyeq 0} \|\mathcal{H}(A) - X\|_2 \ge \min\{r \ge 0 : \ \mathcal{H}(A) + rI \succcurlyeq 0\} . \tag{3.5}$$

Now let $\hat{r} = \min\{r \ge 0 : \ \mathcal{H}(A) + rI \succcurlyeq 0\}$. We have $\mathcal{H}(A) + \hat{r}I \succcurlyeq 0$ and

$$\begin{aligned} \min_{X \succcurlyeq 0} \|\mathcal{H}(A) - X\|_2 &\le \|\mathcal{H}(A) - (\mathcal{H}(A) + \hat{r}I)\|_2 \\ &= \min\{r \ge 0 : \ \mathcal{H}(A) + rI \succcurlyeq 0\} . \end{aligned}$$

The preceding inequality together with (3.5) ensure that (3.4) holds. □

Consider the symmetric/skew-symmetric splittings of the coefficient matrices $A^{(i)} = \mathcal{H}(A^{(i)}) + \mathcal{S}(A^{(i)})$ for $i = 1, 2, \ldots, N$. It is immediate to see that the symmetric and skew-symmetric parts of $\mathcal{A}$ have the forms, respectively,

$$\mathcal{H}(\mathcal{A}) = \sum_{j=1}^{N} I^{(I_N)} \otimes \ldots \otimes I^{(I_{j+1})} \otimes \mathcal{H}(A^{(j)}) \otimes I^{(I_{j-1})} \otimes \ldots \otimes I^{(I_1)}$$

and

$$\mathcal{S}(\mathcal{A}) = \sum_{j=1}^{N} I^{(I_N)} \otimes \ldots \otimes I^{(I_{j+1})} \otimes \mathcal{S}(A^{(j)}) \otimes I^{(I_{j-1})} \otimes \ldots \otimes I^{(I_1)}. \tag{3.6}$$

PROPOSITION 3.6. *Let $\mathcal{A}$ be defined by* (1.6). *The following relation holds for the spectral norm of its skew-symmetric part,*

$$\|\mathcal{S}(\mathcal{A})\|_{2} = \sum_{j=1}^{N} \|\mathcal{S}(A^{(j)})\|_{2}. \tag{3.7}$$

Proof. It is well known that the skew-symmetric matrix $\mathcal{S}(\mathcal{A})$ is a normal matrix and, therefore, it is unitarily diagonalizable. Now, we can conclude the result immediately from (3.6), which provides a relation between eigenvalues of $\mathcal{S}(\mathcal{A})$ and $\mathcal{S}(A^{(i)})$ for $i = 1, 2, \ldots, N$.
□

Our reason for defining distance in terms of the norm $\| \cdot \|_{ss}$, instead of in terms of the spectral or Frobenius norms, is that the quantities $\delta^{\mp}_{ss}(\mathcal{A})$ are easier to compute than $\delta^{\mp}_{2}(\mathcal{A})$ and $\delta^{\mp}_{F}(\mathcal{A})$ when A is a square matrix with Kronecker structure (1.6). This is discussed in the following remark.

Remark 3.7. The distance of a matrix $\mathcal{A}$ with Kronecker structure (1.6) to the set of positive or negative semi-definite matrices may be measured by using $\delta^{\pm}_{2}(\mathcal{A})$, $\delta^{\pm}_{F}(\mathcal{A})$, or $\delta^{\pm}_{ss}(\mathcal{A})$. We obtain from Propositions 3.5 and 3.6 that

$$\delta^{+}_{ss}(\mathcal{A}) = \max\{0, -\lambda_{\min}(\mathcal{H}(\mathcal{A}))\} + \sum_{j=1}^{N} \|\mathcal{S}(A^{(j)})\|_{2}$$

and

$$\delta^{-}_{ss}(\mathcal{A}) = \max\{0, \lambda_{\max}(\mathcal{H}(\mathcal{A}))\} + \sum_{j=1}^{N} \|\mathcal{S}(A^{(j)})\|_{2}.$$

The eigenvalues of a matrix $\mathcal{A}$ with Kronecker structure (1.6) are all possible sums of the form $\lambda^{(1)}_{i_1} + \lambda^{(2)}_{i_2} + \ldots + \lambda^{(N)}_{i_N}$, where $\lambda^{(j)}_{i_j} \in \sigma(A^{(j)})$ and $1 \le i_j \le I_j$ for $j = 1, 2, \ldots, N$. It is more expensive to compute all positive (or negative) eigenvalues of $\mathcal{H}(\mathcal{A})$ than to evaluate the extreme eigenvalues only. It follows that the quantities $\delta^{\pm}_{ss}(\mathcal{A})$ are cheaper to evaluate than $\delta^{\pm}_{F}(\mathcal{A})$.

We also note that $\|\mathcal{S}(\mathcal{A})\|_{F}$ may be much larger than $\|\mathcal{S}(\mathcal{A})\|_{2}$ for $j = 1, 2, \ldots, N$. In fact,

$$\|\mathcal{S}(\mathcal{A})\|_{F}^{2} = \sum_{j=1}^{N} I_1 \times \ldots \times I_{j-1} \times I_{j+1} \times \ldots \times I_N \times \|\mathcal{S}(A^{(j)})\|_{F}^{2},$$

where $I_0 = 1$ and $A^{(j)} \in \mathbb{R}^{I_j \times I_j}$ for $j = 1, 2, \ldots, N$.

For completeness, we note that

$$\begin{aligned} \|\mathcal{H}(\mathcal{A})\|_{2} &= \max\{|\lambda_{\max}(\mathcal{H}(\mathcal{A}))|, |\lambda_{\min}(\mathcal{H}(\mathcal{A}))|\} \\ &= \max\left\{ \left| \sum_{i=1}^{N} \lambda_{\min}(\mathcal{H}(A^{(i)})) \right|, \left| \sum_{i=1}^{N} \lambda_{\max}(\mathcal{H}(A^{(i)})) \right| \right\}. \end{aligned}$$

Table 2
Distances to the symmetric positive definite and negative definite tensors for Example 3.8.

n	$\Vert\mathcal{S}(\mathcal{A})\Vert_2$	$\Vert\mathcal{H}(\mathcal{A})\Vert_2$	$\frac{\Vert\mathcal{S}(\mathcal{A})\Vert_2}{\Vert\mathcal{A}\Vert_{ss}}$	$\frac{\Vert\mathcal{H}(\mathcal{A})\Vert_2}{\Vert\mathcal{A}\Vert_{ss}}$	$\frac{\delta^{+}_{ss}(\mathcal{A})}{\Vert\mathcal{A}\Vert_{ss}}$	$\frac{\delta^{-}_{ss}(\mathcal{A})}{\Vert\mathcal{A}\Vert_{ss}}$
100	2.998	2.998	0.5	0.5	1	1
500	2.999	2.999	0.5	0.5	1	1
1000	3	3	0.5	0.5	1	1

We turn to a simple test example for which the GMRES_BTF algorithm [11] breaks down after a few steps and produces a poor approximate solution of the Sylvester tensor equation that we try to solve. It can be shown that the use of GMRES_BTF is mathematically equivalent to the application of GMRES to the solution of the linear system of equations (1.5); see [5]. To be able to discuss the cause of breakdown in a simple manner, we consider the equivalent linear system of equations (1.5).

Example 3.8. *Consider the Sylvester tensor equation*

$$\mathcal{X} \times_1 A^{(1)} + \mathcal{X} \times_2 A^{(2)} + \mathcal{X} \times_3 A^{(3)} = \mathcal{D}, \tag{3.8}$$

in which the $A^{(i)}$, for $i = 1, 2, 3$, are $n \times n$ downshift matrices, i.e,

$$A^{(i)} = \begin{bmatrix} 0 & 0 & 0 & \dots & 0 & 0 \\ 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & 0 & \dots & 0 & 0 \\ & 0 & \ddots & \vdots & 0 & 0 \\ & & \ddots & & 0 & 0 \\ & & & 0 & 1 & 0 \end{bmatrix} \in \mathbb{R}^{n\times n}. \tag{3.9}$$

Regard the linear system (1.5) *corresponding to* (3.8)*, and choose the right-hand side $\mathcal{D}$ so that $vec(\mathcal{D}) = e_n \otimes e_{n-m+1} \otimes e_n$ for some integer $1 \leq m < n$. Then $\mathcal{X}^*$ with $vec(\mathcal{X}^*) = e_n \otimes e_{n-m} \otimes e_n$ is a solution of* (3.8)*. The GMRES_BTF algorithm [11] applied to the solution of* (3.8) *with the zero tensor as initial iterate breaks down at step m. In particular, for $m = 1$, the GMRES_BTF algorithm determines the zero tensor as approximate solution when it breaks down at the first step. Thus, GMRES_BTF is not a useful solution method for this problem. Related examples when the linear system of equations* (1.5) *does not have a tensor structure and the matrix is of the form* (3.9) *are discussed in [10, 18]. We remark that while the system* (3.8) *is artificial, related systems are obtained when seeking to deblur color images that have been contaminated by noise and motion blur. A discussion on the deblurring of monochromatic images that have been contaminated by noise and motion blur can be found in [13]. Table 2 reports the distance of $\mathcal{A}$ in* (1.5) *to the sets of (skew-)symmetric and positive (negative) semi-definite matrices.*

The relative distance of $\mathcal{A}$ in the present example to the set of symmetric (symmetric positive semi-definite) matrices is equal to the distance to the set of skew-symmetric (symmetric negative semi-definite) matrices. This is shown in [18] in the situation when $\mathcal{A}$ only consists of the matrix (3.9)*. The proof can be adapted to the present situation. The tensor $\mathcal{A}$ may be considered the "worst" tensor for GMRES_BTF.*

We conclude this section with an example that includes different matrices in the Kronecker structure (1.6), but their distances to the sets of symmetric matrices and skew-symmetric matrices are almost equal. The GMRES_BTF algorithm is seen to perform quite differently for one of the mentioned cases.

Example 3.9. *Let $N = 2$, $n_1 = 500$, and $n_2 = 2$. We consider the solution of a Sylvester tensor equations using the GMRES_BTF algorithm with the zero tensor as initial* *iterate. The equation is described by*

$$\mathcal{X} \times_1 \tilde{A}^{(1)} + \mathcal{X} \times_2 A^{(2)} = \mathcal{D}, \tag{3.10}$$

with $\tilde{A}^{(1)} = A^{(1)} + \mathrm{tridiag}(-\alpha, 0, \alpha)$, *where the diagonal and off-diagonal entries of* $A^{(1)}$ *are equal to* 8 *and* 5, *respectively. Here* α *is a prescribed nonnegative parameter and*

$$A^{(2)} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \in \mathbb{R}^{2\times 2}$$

is the downshift matrix. The right-hand side tensor $\mathcal{D}$ *is constructed so that* $\mathcal{X}^* \in \mathbb{R}^{n_1 \times 2 \times n_1}$, *with* $vec(\mathcal{X}^*) = e_{n_1} \otimes e_1 \otimes e_{n_1}$, *solves the Sylvester tensor equation* (3.10).

Let $\mathcal{A}$ *be the matrix in form* (1.6) *that corresponds to* (3.10). *We applied the GMRES_BTF algorithm (without restarting) using an implementation based on Givens rotations. The iterations were terminated as soon as the relative norm of the residuals became less than* 10^{-11}. *We remark that residual tensors were not formed explicitly and their norm were computed by using Givens rotations; see [41, Chapter 6].*

Table 3 shows the performance of the GMRES_BTF method, as well as the relative distances of the Kronecker structure (1.6) *to the set of all symmetric and skew-symmetric matrices for different values of* α. *The matrix* $\tilde{\mathcal{X}}$ *denotes the computed approximate solution, and "Iter" stands for the required number of iterations to satisfy the stopping criterion of the algorithm. The GMRES_BTF method can be seen to terminate after four iterations and determines a poor approximate solution when* $\alpha \leq 10^{-9}$.

Notice that for $\alpha \leq 0.001$, *the relative distances to the set of symmetric matrices are the same to four decimal digits. However, the GMRES_BTF algorithm performs much better when* α *is not too small. This example illustrates that the performance of GMRES_BTF does not only depend on the distance of the matrix* (1.5) *to the set of symmetric matrices and the set of skew-symmetric matrices. It is well-known that the performance of GMRES when applied to the solution of a linear system of equations with a square nonsingular matrix depends on the eigenvalues and eigenvectors of the matrix, as well as on the right-hand side; see Du et al. [14] for a recent discussion.*

Table 3

Relative distances of $\mathcal{A}$ *in* (1.5) *to the set of (skew-)symmetric matrices and the performance of GMRES_BTF for* (3.10).

α	0	10^{-9}	10^{-6}	10^{-3}	10^{-1}	1	10
$\Vert\mathcal{S}(\mathcal{A})\Vert_2$	0.5000	0.5000	0.5000	0.5020	0.7000	2.5000	20.4996
$\Vert\mathcal{H}(\mathcal{A})\Vert_2$	$2.5035 \cdot 10^3$	$2.5035 \cdot 10^3$	$2.5035 \cdot 10^3$	$2.5035 \cdot 10^3$	$2.5035 \cdot 10^3$	$2.5035 \cdot 10^3$	$2.5035 \cdot 10^3$
$\frac{\Vert\mathcal{S}(\mathcal{A})\Vert_2}{\Vert\mathcal{A}\Vert_{ss}}$	$1.9968 \cdot 10^{-4}$	$1.9968 \cdot 10^{-4}$	$1.9968 \cdot 10^{-4}$	$2.0048 \cdot 10^{-4}$	$2.7953 \cdot 10^{-4}$	$9.9759 \cdot 10^{-4}$	0.0081
$\frac{\Vert\mathcal{H}(\mathcal{A})\Vert_2}{\Vert\mathcal{A}\Vert_{ss}}$	0.9998	0.9998	0.9998	0.9998	0.9997	0.9990	0.9919
$\frac{\Vert\tilde{\mathcal{X}}-\mathcal{X}^*\Vert}{\Vert\mathcal{X}^*\Vert}$	0.3151	0.3151	$2.2147 \cdot 10^{-7}$	$1.1073 \cdot 10^{-7}$	$3.1635 \cdot 10^{-9}$	$7.9453 \cdot 10^{-10}$	$2.8813 \cdot 10^{-10}$
Iter	4	4	5	6	10	23	165

4. Tikhonov regularization methods based on tensor format.

We present several iterative schemes in tensor framework. Two of these methods apply the Arnoldi and generalized Arnoldi processes to the approximate solution of the Tikhonov minimization problem (1.4). The corresponding algorithms are referred to as the Arnoldi–Tikhonov method based on tensor format (AT_BTF) and the generalized AT_BTF (GAT_BTF) method. These algorithms generalize methods discussed by Huang et al. [24] for the situation when there is no exploitable tensor structure; see also Bouhamidi et al. [8]. In the situation when all matrices $A^{(i)}$ are symmetric, the AT_BTF method reduces to the Lanczos–Tikhonov method based on tensor format (LT_BTF). In addition, we describe a flexible AT_BTF method, which will be referred to as FAT_BTF. This method is an adaption of the flexible Arnoldi process introduced by Saad [41], and more recently discussed by Gazzola and Nagy [19] and Morikuni et al. [37], for the solution of linear systems of equations with no tensor structure, to the solution of Sylvester tensor equations. Details of the derivations of the algorithm are left to the reader.

4.1. The AT_BTF method. Introduce the linear operator

$$\begin{array}{rcl} \mathcal{M} : \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N} & \to & \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}, \\ \mathcal{X} & \mapsto & \mathcal{M}(\mathcal{X}) := \mathcal{X} \times_1 A^{(1)} + \mathcal{X} \times_2 A^{(2)} + \ldots + \mathcal{X} \times_N A^{(N)} \end{array} \tag{4.1}$$

and define the tensor Krylov subspace

$$\mathcal{K}_m(\mathcal{M}, \mathcal{D}) = \text{span}\{\mathcal{D}, \mathcal{M}(\mathcal{D}), \ldots, \mathcal{M}^{m-1}(\mathcal{D})\}.$$

The Arnoldi_BTF process, described by Algorithm 1, produces an orthonormal basis for $\mathcal{K}_m(\mathcal{M}, \mathcal{D})$, provided that m is small enough so that breakdown does not occur. Let $\tilde{\mathcal{V}}_\ell$ denote the $(N+1)$-mode tensor with the column tensors $\mathcal{V}_1, \mathcal{V}_2, \ldots, \mathcal{V}_\ell$ for $1 \leq \ell \leq m+1$ produced by Algorithm 1. It is not difficult to see that $\tilde{\mathcal{V}}_\ell \boxtimes^{(N+1)} \tilde{\mathcal{V}}_\ell = I_\ell$ for $1 \leq \ell \leq m$. In the sequel, the matrix $\bar{H}_m = [h_{ij}] \in \mathbb{R}^{(m+1) \times m}$ is of upper Hessenberg form; its nontrivial entries are computed in lines 6 and 9 of Algorithm 1. The matrix $H_m \in \mathbb{R}^{m \times m}$ is obtained by deleting the last row of $\bar{H}_m$. It is shown in [5, Theorem 3.1] that

$$\tilde{\mathcal{W}}_m = \tilde{\mathcal{V}}_{m+1} \times_{(N+1)} \bar{H}_m^T, \tag{4.2}$$

where $\tilde{\mathcal{W}}_m$ is an $(N+1)$-mode tensor generated by Algorithm 1, has the column tensors $\mathcal{W}_j := \mathcal{M}(\mathcal{V}_j)$ for $j = 1, 2, \ldots, m$. Moreover, by [5, Proposition 3.2], we have

$$\tilde{\mathcal{V}}_m \boxtimes^{(N+1)} \tilde{\mathcal{W}}_m = H_m \tag{4.3}$$

and $\tilde{\mathcal{V}}_{m+1} \boxtimes^{(N+1)} \tilde{\mathcal{W}}_m = \bar{H}_m$.

Algorithm 1: The Arnoldi_BTF process [5].

```
1  Input: Coefficient matrices $A^{(1)}, A^{(2)}, \ldots, A^{(N)}$, right-hand side $\mathcal{D}$, and initial
     approximate solution $\mathcal{X}_0$.
2  Compute $\mathcal{R}_0 := \mathcal{D} - \mathcal{M}(\mathcal{X}_0)$, $\beta := \|\mathcal{R}_0\|$, $\mathcal{V}_1 := \mathcal{R}_0/\beta$.
3  for $j = 1, 2, \ldots, m$ do
4      $\mathcal{W} := \mathcal{M}(\mathcal{V}_j)$;
5      for $i = 1, 2, \ldots, j$ do
6          $h_{ij} := \langle \mathcal{W}, \mathcal{V}_i \rangle$ ;
7          $\mathcal{W} := \mathcal{W} - h_{ij}\mathcal{V}_i$;
8      end
9      $h_{j+1,j} := \|\mathcal{W}\|$. If $h_{j+1,j} := 0$, then stop;
10     $\mathcal{V}_{j+1} := \mathcal{W}/h_{j+1,j}$;
11 end
```

If the matrices $A^{(1)}, A^{(2)}, \ldots, A^{(N)}$ are symmetric positive definite, then in view of (4.3), the Hessenberg matrix H_m is symmetric positive definite and, therefore, tridiagonal. It follows that Algorithm 1 reduces to the Lanczos process based on tensor format (Lanczos_BTF), which is described in [5, Algorithm 2].

Having carried out $m := k$ steps with Algorithm 1, we determine an approximate solution $\mathcal{X}_k \in \mathcal{K}_k(\mathcal{M}, \mathcal{D})$ of (1.4) as follows. Let

$$\mathcal{X}_k = \sum_{i=1}^{k} y_k^{(i)} \mathcal{V}_i = \tilde{\mathcal{V}}_k \bar{\times}_{(N+1)} y_k, \quad \text{where} \;\; y_k = (y_k^{(1)}; \ldots ; y_k^{(k)}) \in \mathbb{R}^k.$$

Using (4.2) and Proposition 1.1, we obtain

$$\|\mathcal{D} - \mathcal{M}(\mathcal{X}_k)\| = \| \, \|\mathcal{D}\| \, e_1 - \bar{H}_k y_k \|_2$$

and

$$\left\| \sum_{j=1}^{M} \mathcal{X}_k \times_j L^{(j)} \right\|^2 = y_k^T \left(\tilde{\mathcal{M}}_k \boxtimes^{(N+1)} \tilde{\mathcal{M}}_k \right) y_k,$$

where

$$\tilde{\mathcal{M}}_k = \sum_{j=1}^{M} \tilde{\mathcal{V}}_k \times_j L^{(j)}$$

is an $(N+1)$-mode tensor. It follows from these relations that (1.4) can be expressed as the Tikhonov minimization problem

$$\min_{y \in \mathbb{R}^k} \left\{ \left\| \|\mathcal{D}\| \, e_1 - \bar{H}_k y \right\|_2^2 + \lambda y^T \left(\tilde{\mathcal{M}}_k \boxtimes^{(N+1)} \tilde{\mathcal{M}}_k \right) y \right\} \tag{4.4}$$

with solution $y = y_k \in \mathbb{R}^k$.

The expression (4.4) generalizes a solution method proposed by Huang et al. [24] to equations (1.4) with a tensor structure. The matrix

$$\mathcal{N}_k = \tilde{\mathcal{M}}_k \boxtimes^{(N+1)} \tilde{\mathcal{M}}_k \tag{4.5}$$

in the regularization term is a Gram matrix and, therefore, positive semi-definite.

Referring to [24, eq. (3.6)], without going into details, let N_k denote the Gram matrix there (which corresponds to the matrix $\mathcal{N}_k$ introduced above). The quadratic form $y^T N_k y$ in [24, eq. (3.6)] is replaced by $\|\tilde{L}_k^T y\|_2^2$, where the matrix $\tilde{L}_k$ has to be determined. When N_k is positive definite, Huang et al. [24] let $\tilde{L}_k$ be the Cholesky factor of N_k; when N_k is singular they propose to let $\tilde{L}_k^T = D_k^{1/2} Q_k^T$, where $N_k = Q_k D_k Q_k^T$, $Q_k \in \mathbb{R}^{k \times k}$ is an orthogonal matrix and the matrix D_k is diagonal. We follow the same strategy for the Gram matrix (4.5) and obtain from (4.4) a Tikhonov minimization problem in general form,

$$\min_{y \in \mathbb{R}^k} \left\{ \left\| \|\mathcal{D}\| \, e_1 - \bar{H}_k y \right\|_2^2 + \lambda \left\| \tilde{L}_k^T y \right\|_2^2 \right\}. \tag{4.6}$$

In applications of interest to us, k generally is fairly small. We therefore may solve (4.6) by computing the generalized singular value decomposition of the matrix pair $\{\bar{H}_k, \tilde{L}_k^T\}$; see, e.g., [15, 20]. Another solution approach is to let $\hat{H}_k = \bar{H}_k \tilde{L}_k^{-T}$ and express (4.6) as a Tikhonov minimization problem in standard form,

$$\min_{z \in \mathbb{R}^k} \left\{ \left\| \|\mathcal{D}\| \, e_1 - \hat{H}_k z \right\|_2^2 + \lambda \, \|z\|_2^2 \right\}. \tag{4.7}$$

This approach is discussed in [24, cf. (3.8) and (3.9)]. Since in many applications the matrix $\tilde{L}_k$ is not very ill-conditioned, the solution of linear systems of equations with the matrix $\tilde{L}_k$, which is required when forming $\hat{H}_k$, is feasible.

It is interesting to investigate when the matrix $\mathcal{N}_k$ is invertible. The following result provides sufficient conditions. Using a similar technique, one can derive sufficient conditions for nonsingularity of the matrix N_k in [24, Eq. (3.6)].

THEOREM 4.1. *Assume that k steps of Algorithm 1 have been carried out and let $\tilde{\mathcal{V}}_k$ be the $(N+1)$-mode tensor with the column tensors $\mathcal{V}_1, \mathcal{V}_2, \ldots, \mathcal{V}_k$ determined by the algorithm. Let $\tilde{\mathcal{M}}_k = \sum_{j=1}^{M} \tilde{\mathcal{V}}_k \times_j L^{(j)}$. If the matrix*

$$\mathcal{L} := \sum_{j=1}^{M} I^{(I_M)} \otimes \ldots \otimes I^{(I_{j+1})} \otimes L^{(j)} \otimes I^{(I_{j-1})} \otimes \ldots \otimes I^{(I_1)} \tag{4.8}$$

is invertible, or if

$$Null(\mathcal{L}) \cap span\{vec(\mathcal{V}_1), \ldots, vec(\mathcal{V}_k)\} = \{0\},$$

then the $k \times k$ matrix $\mathcal{N}_k = \tilde{\mathcal{M}}_k \boxtimes^{(N+1)} \tilde{\mathcal{M}}_k$ is nonsingular.

Proof. Since $\mathcal{N}_k \in \mathbb{R}^{k \times k}$ is a Gram matrix, we only need to show that the frontal slices of $\tilde{\mathcal{M}}_k$ are linearly independent. This ensures the invertibility of $\mathcal{N}_k$. As $M \leq N$, we conclude that k frontal slices of the $(N+1)$-mode tensor $\tilde{\mathcal{M}}_k$ are given by

$$\mathcal{M}_\ell = \sum_{j=1}^{M} \mathcal{V}_\ell \times_j L^{(j)}, \quad \ell = 1, 2, \ldots, k.$$

Suppose that

$$0 = \sum_{\ell=1}^{k} \alpha_\ell \mathcal{M}_\ell = \sum_{\ell=1}^{k} \sum_{j=1}^{M} \alpha_\ell \mathcal{V}_\ell \times_j L^{(j)} = \sum_{j=1}^{M} \left(\sum_{\ell=1}^{k} \alpha_\ell \mathcal{V}_\ell \right) \times_j L^{(j)}$$

for some scalars $\alpha_1, \alpha_2, \ldots, \alpha_k$. The above relation is equivalent to

$$\sum_{\ell=1}^{k} \alpha_\ell \text{vec}(\mathcal{V}_\ell) \in \text{Null}\,(\mathcal{L}).$$

The vectors $\text{vec}(\mathcal{V}_1), \text{vec}(\mathcal{V}_2), \ldots, \text{vec}(\mathcal{V}_k)$ are linearly independent. The assertion therefore follows. □

Remark 4.2. The spectrum of $\mathcal{L}$ is given by

$$\sigma(\mathcal{L}) = \left\{ \sum_{i=1}^{M} \lambda_i \ : \ \lambda_i \in \sigma(L^{(i)}) \quad \text{for} \quad i = 1, 2, \ldots, M \right\}.$$

This shows that if all regularization matrices $L^{(i)}$ are symmetric positive semi-definite, with at least one of them positive definite, then $\mathcal{L}$ is invertible.

The reduced Tikhonov minimization problem in standard form (4.7) is solved by the technique used in [24], where the regularization parameter λ is determined with the aid of the discrepancy principle (see Appendix A for further details). Algorithm 2 summarizes the computations. The algorithm can be implemented by using the MATLAB Tensor Toolbox [2].

4.2. The flexible AT_BTF method. We describe how the flexible Arnoldi process discussed in [19, 37, 41] can be adapted to the tensor framework. We refer to the iterative scheme as the flexible Arnoldi method based on tensor format (FAT_BTF). The computations are summarized in Algorithm 3. We remark that we may replace line 4 of the algorithm by other ways for determining a suitable basis $\mathcal{Z}_1, \mathcal{Z}_2, \ldots, \mathcal{Z}_m$ of the solution subspace. For

Algorithm 2: The AT_BTF (LT_BTF) regularization method.

```
Input: The coefficient matrices $A^{(i)}$, $i = 1, 2, \dots, N$; the right-hand side tensor $\mathcal{D}$;
  the regularization matrices $L^{(j)}$, $j = 1, 2, \dots, M$, chosen so that (4.8) is
  nonsingular (cf. Remark 4.2); and the parameters $\delta$ and $\eta > 1$ used in the
  discrepancy principle (see Appendix A for details).
for $k = 1, 2, \dots$ until convergence do
    Compute $\tilde{\mathcal{V}}_k$ with the column tensors $\mathcal{V}_i$ and $\bar{H}_k$ by Algorithm 1 (or by
      Lanczos_BTF process if all the $A^{(i)}$ are symmetric);
    Compute column tensors of $\tilde{\mathcal{M}}_k$ by $\mathcal{M}_i = \sum_{j=1}^{M} \mathcal{V}_i \times_j L^{(j)}$ $(i = 1, 2, \dots, k)$;
    Compute $\mathcal{N}_k = \tilde{\mathcal{M}}_k \boxtimes^{(N+1)} \tilde{\mathcal{M}}_k$;
    Compute the Cholesky factorization of $\mathcal{N}_k = \tilde{L}_k \tilde{L}_k^T$;
    Compute $\hat{H}_k = \bar{H}_k \tilde{L}_k^{-T}$;
    Compute the zero $\lambda > 0$ of $\phi(\lambda) = \| \|\mathcal{D}\| e_1 - \hat{H}_k z_{\lambda,k} \|_2^2 - \eta^2 \delta^2$, where $\delta$ is a
      bound for the norm of the error $\mathcal{E}$ in $\mathcal{D}$. We comment that the vector $z_{\lambda,k}$ in
      $\phi(\lambda)$ is written as a (one-variable) function of $\lambda$; see Appendix A for more
      details. After computing the regularization parameter $\lambda$, determine the vector
      $z_{\lambda,k}$ by solving the following (small) least-squares problem

          $\min_{z \in \mathbb{R}^k} \left\| \begin{bmatrix} \hat{H}_k \\ \lambda^{1/2} I_k \end{bmatrix} z - \begin{bmatrix} \|\mathcal{D}\| e_1 \\ 0 \end{bmatrix} \right\|_2^2;$

      Let $y_k = (y_k^{(1)}; \dots; y_k^{(k)})^T := \tilde{L}_k^{-T} z_{\lambda,k}$ and compute
      $\mathcal{X} = \sum_{i=1}^{k} y_k^{(i)} \mathcal{V}_i = \tilde{\mathcal{V}}_k \bar{\times}_{(N+1)} y_k$;
end
```

instance, we may carry out more steps with the BiCGSTAB_BTF method [11]. We found the basis determined by Algorithm 3 to perform well for the problems of Section 5.

Having carried out k steps with Algorithm 3, we can determine an approximate solution $\mathcal{X}_k$ of (1.4) of the form

$$\mathcal{X}_k = \sum_{i=1}^{k} y_k^{(i)} \mathcal{Z}_i = \tilde{\mathcal{Z}}_k \bar{\times}_{(N+1)} y_k, \qquad y_k = (y_k^{(1)}; \dots; y_k^{(k)}) \in \mathbb{R}^k, \tag{4.9}$$

in which $\tilde{\mathcal{Z}}_k$ denotes the $(N+1)$-mode tensor with the column tensors $\mathcal{Z}_1, \mathcal{Z}_2, \dots, \mathcal{Z}_k$. It is not difficult to see that

$$\mathcal{D} - \mathcal{M}(\mathcal{X}_k) = \tilde{\mathcal{V}}_k \bar{\times}_{(N+1)} (\|\mathcal{D}\|\, e_1 - \bar{H}_k y_k)$$

and

$$\left\| \sum_{j=1}^{M} \mathcal{X}_k \times_j L^{(j)} \right\|^2 = y_k^T \left(\hat{\mathcal{M}}_k \boxtimes^{(N+1)} \hat{\mathcal{M}}_k \right) y_k,$$

where $\hat{\mathcal{M}}_k = \sum_{j=1}^{M} \tilde{\mathcal{Z}}_k \times_j L^{(j)}$, and $\tilde{\mathcal{Z}}_k$ denotes the $(N+1)$-mode tensor with the column tensors $\mathcal{Z}_1, \mathcal{Z}_2, \dots, \mathcal{Z}_k$. Using the above relations, the Tikhonov minimization problem (1.4) can be

Algorithm 3: The flexible Arnoldi_BTF process.

```
Input: Coefficient matrices $A^{(1)}, A^{(2)}, \ldots, A^{(N)}$, right-hand side $\mathcal{D}$, and initial
  approximate solution $\mathcal{X}_0$.
Compute $\mathcal{R}_0 = \mathcal{D} - \mathcal{M}(\mathcal{X}_0)$, $\beta := \|\mathcal{R}_0\|$, $\mathcal{V}_1 := \mathcal{R}_0/\beta$, $h_{0,1} = 0$, and $\mathcal{V}_0 = 0$.
for $j = 1, 2, \ldots, m$ do
    Apply two steps of BiCGSTAB_BTF [11] to find $\mathcal{Z}_j$ as an approximate solution
      of $\mathcal{M}(\mathcal{Z}_j) = \mathcal{V}_j$;
    $\mathcal{W} := \mathcal{M}(\mathcal{Z}_j)$;
    for $i = 1, 2, \ldots, j$ do
        $h_{ij} := \langle \mathcal{W}, \mathcal{V}_i \rangle$ ;
        $\mathcal{W} := \mathcal{W} - h_{ij}\mathcal{V}_i$;
    end
    $h_{j+1,j} := \|\mathcal{W}\|$. If $h_{j+1,j} := 0$, then stop;
    $\mathcal{V}_{j+1} := \mathcal{W}/h_{j+1,j}$;
end
```

reduced to

$$\min_{y\in\mathbb{R}^k}\left\{\left\|\|\mathcal{D}\|\,e_1 - \bar{H}_k y\right\|_2^2 + \lambda y^T\left(\hat{\mathcal{M}}_k \boxtimes^{(N+1)} \hat{\mathcal{M}}_k\right) y\right\}. \tag{4.10}$$

This minimization problem can be solved similarly as the corresponding minimization problem of the previous subsection. We refer to this solution method as the flexible Arnoldi–Tikhonov method based on tensor format. We will abbreviate it by FAT_BTF also.

4.3. The GAT_BTF method. Consider the Tikhonov regularization problem

$$\min_{x\in\mathbb{R}^n}\left\{\|Ax-b\|_2^2 + \lambda\,\|Lx\|_2^2\right\},$$

where A is a severely ill-conditioned matrix and L is a general regularization matrix. A technique for determining an approximate solution in a generalized Krylov subspace is described in [38]. This method is based on simultaneously reducing the matrices A and L by a generalized Arnoldi process proposed by Li and Ye [32]. The method is extended in [8] to the solution of a class ill-posed matrix equation. Here we describe an extension that can be applied to the solution of equation (1.4). We refer to the resulting scheme as the GAT_BTF method.

Introduce the linear operator

$$\begin{array}{rcl} \mathcal{L} : \mathbb{R}^{I_1\times I_2\times\ldots\times I_N} & \to & \mathbb{R}^{I_1\times I_2\times\ldots\times I_N}, \\ \mathcal{X} & \mapsto & \mathcal{L}(\mathcal{X}) := \mathcal{X}\times_1 L^{(1)} + \mathcal{X}\times_2 L^{(2)} + \ldots + \mathcal{X}\times_M L^{(M)}. \end{array} \tag{4.11}$$

Algorithm 4, which generalizes the method described in [38], generates generalized Krylov subspaces spanned by elements of the form

$$\mathcal{D}, \mathcal{M}(\mathcal{D}), \mathcal{L}(\mathcal{D}), \mathcal{M}^2(\mathcal{D}), \mathcal{M}(\mathcal{L}(\mathcal{D})), \mathcal{L}(\mathcal{M}(\mathcal{D})), \mathcal{L}^2(\mathcal{D}), \ldots\ . \tag{4.12}$$

The execution of k steps of this algorithm requires k applications of the operators (4.1) and (4.11). A generalization of the algorithm in [38] is described in [40]. The latter algorithm also can be modified to be applicable to the operators (4.1) and (4.11).

Let α_k and β_k stand for the values of ℓ in lines 17 and 27, respectively, of Algorithm 4, and let $\bar{\mathcal{H}}_{\mathcal{M},k} = [\mathcal{H}_{\mathcal{M}}(i,j)] \in \mathbb{R}^{\alpha_k\times k}$ and $\bar{\mathcal{H}}_{\mathcal{L},k} = [\mathcal{H}_{\mathcal{L}}(i,j)] \in \mathbb{R}^{\beta_k\times k}$ denote the matrices, whose nontrivial entries are computed in lines 10 and 13, and in lines 20 and 23, respectively, of the algorithm. In the sequel, suppose that $\tilde{\mathcal{V}}_{\alpha_k}$ and $\tilde{\mathcal{V}}_{\beta_k}$ are $(N+1)$-mode tensors, whose column tensors $\mathcal{V}_i$ are determined by Algorithm 4. It is not difficult to see that $\tilde{\mathcal{V}}_{\alpha_k} \boxtimes^{(N+1)} \tilde{\mathcal{V}}_{\alpha_k} = I^{(\alpha_k)}$ and $\tilde{\mathcal{V}}_{\beta_k} \boxtimes^{(N+1)} \tilde{\mathcal{V}}_{\beta_k} = I^{(\beta_k)}$. With a strategy similar to the one used in [5, Theorem 3.1], it can be shown that

$$\tilde{\mathcal{W}} = \tilde{\mathcal{V}}_{\alpha_k} \times_{(N+1)} \bar{\mathcal{H}}_{\mathcal{M},k}^T \quad \text{and} \quad \hat{\mathcal{W}} = \tilde{\mathcal{V}}_{\beta_k} \times_{(N+1)} \bar{\mathcal{H}}_{\mathcal{L},k}^T, \tag{4.13}$$

where $\tilde{\mathcal{W}}$ and $\hat{\mathcal{W}}$ are two $(N+1)$-mode tensors with column tensors $\tilde{\mathcal{W}}_j := \mathcal{M}(\mathcal{V}_j)$ and $\hat{\mathcal{W}}_j := \mathcal{L}(\mathcal{V}_j)$ for $j = 1, 2, \ldots, k$

The following proposition is useful for deriving a Tikhonov regularization problem of low dimension by projecting (1.4) onto a generalized Krylov subspace spanned by elements of the form (4.12). The proof follows from properties of contracted product and straightforward computations. We omit the details.

Proposition 4.3. *Let $\tilde{\mathcal{V}}_r$ be an $(N+1)$-mode tensor with the column tensors $\mathcal{V}_j$, for $j = 1, 2, \ldots, r$, such that $\tilde{\mathcal{V}}_r \boxtimes^{(N+1)} \tilde{\mathcal{V}}_r = I^{(r)}$. Assume that $\tilde{\mathcal{D}}_r = \tilde{\mathcal{V}}_r \bar{\times}_{(N+1)} z_{\mathcal{D}}$ with $z_{\mathcal{D}} = \tilde{\mathcal{V}}_r \boxtimes^{(N+1)} \mathcal{D} = \left(\langle \mathcal{V}_1, \mathcal{D}\rangle, \ldots, \langle \mathcal{V}_r, \mathcal{D}\rangle\right)^T$, where $\mathcal{D}$ is an N-mode tensor. For all $z, d \in \mathbb{R}^r$, we have*

1. *$\langle \tilde{\mathcal{V}}_r \bar{\times}_{(N+1)} z, \tilde{\mathcal{V}}_r \bar{\times}_{(N+1)} d\rangle = \langle z, d\rangle_2$ and $\|\tilde{\mathcal{V}}_r \bar{\times}_{(N+1)} z\| = \|z\|_2$,*
2. *$\langle \tilde{\mathcal{V}}_r \bar{\times}_{(N+1)} z, \mathcal{D}\rangle = \langle z, z_{\mathcal{D}}\rangle_2$,*
3. *$\|\mathcal{D} - \tilde{\mathcal{D}}_r\|^2 = \|\mathcal{D}\|^2 - \|z_{\mathcal{D}}\|_2^2$,*
4. *$\|\tilde{\mathcal{V}}_r \bar{\times}_{(N+1)} z - \mathcal{D}\|^2 = \|z - z_{\mathcal{D}}\|_2^2 + \|\mathcal{D}\|^2 - \|z_{\mathcal{D}}\|_2^2$,*
5. *if $\mathcal{V}_1 = \mathcal{D}/\|\mathcal{D}\|$, then $\|\tilde{\mathcal{V}}_r \bar{\times}_{(N+1)} z - \mathcal{D}\|^2 = \|z - \|\mathcal{D}\| e_1\|_2^2$.*

Here $\langle x, y\rangle_2$ denotes the standard Euclidean inner product between two real vectors x and y of the same size, i.e., $\langle x, y\rangle_2 = x^T y$.

Consider the subspaces $\mathcal{F}_k = \operatorname{span}\{\mathcal{V}_1, \mathcal{V}_2, \ldots, \mathcal{V}_k\}$ for $k = 0, 1, 2, \ldots$ and let $\tilde{\mathcal{V}}_r$ be an $(N+1)$-mode tensor with the column tensors $\mathcal{V}_i$, for $i = 1, 2, \ldots, r$, generated by Algorithm 4. After k steps of the algorithm, the GAT_BTF method determines an approximate solution $\mathcal{X}_k \in \mathcal{F}_k$ of the form

$$\mathcal{X}_k = \sum_{i=1}^{k} y_k^{(i)} \mathcal{V}_i = \tilde{\mathcal{V}}_k \bar{\times}_{(N+1)} y_k, \qquad y_k = (y_k^{(1)}, \ldots, y_k^{(k)})^T.$$

Using (4.13) and Proposition 4.3, one may easily verify that equation (1.4) can be reduced to the low-dimensional problem,

$$\min_{y_k \in \mathbb{R}^k} \left\{ \|\bar{\mathcal{H}}_{\mathcal{M},k} y_k - \|\mathcal{D}\| e_1\|_2^2 + \lambda \|\bar{\mathcal{H}}_{\mathcal{L},k} y_k\|_2^2 \right\},$$

which can be solved by one of the techniques described in Subsection 4.1.

We conclude this section by noting that all of the iterative schemes described in this section can be used to solve operator equations of the form $\mathcal{S}(\mathcal{X}) = \mathcal{C}$, where $\mathcal{S}(\cdot)$ is a fairly general linear operator from $\mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$ to $\mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$.

5. Numerical experiments. We report results for three test problems to illustrate the performance of the algorithms described. The right-hand side tensor $\mathcal{D}$ in (1.1) is contaminated by an error tensor $\mathcal{E}$ with normally distributed random entries with zero mean and scaled to correspond to a specific noise level $\nu := \|\mathcal{E}\|/\|\mathcal{D}\|$. All computations were carried out on a computer with an Intel Core i7-4770K CPU @ 3.50GHz processor and 24GB RAM using MATLAB R2014a. We used the Tensor Toolbox [2].

The first example stems from the discretization of a dimensionless radiative transfer equation (RTE) by a spectral method. This gives rise to an equation of the form (1.1) with

Algorithm 4: The GAT_BTF process for operator pairs $\{\mathcal{M}, \mathcal{L}\}$.

```
Input: the coefficient $A^{(i)}, i = 1, 2, \dots, N$; the right-hand side tensor $\mathcal{D}$; the
  parameter associated with the discrepancy principle $\eta > 1$ (see Appendix A for
  more details); and an integer $k > 0$.
Choose the regularization matrices $L^{(j)}$s for $j = 1, 2, \dots, M$;
Set $\ell = 1$ and $\mathcal{V}_1 := \mathcal{D}/\|\mathcal{D}\|$;
for $j = 1, 2, \dots, k$ do
    if $j > \ell$ then
        exit;
    end
    $\mathcal{W} := \mathcal{M}(\mathcal{V}_j)$;
    for $i = 1, \dots, \ell$ do
        $\mathcal{H}_{\mathcal{M}}(i, j) := \langle \mathcal{W}, \mathcal{V}_i \rangle$;
        $\mathcal{W} = \mathcal{W} - \mathcal{H}_{\mathcal{M}}(i, j)\mathcal{V}_i$;
    end
    $\mathcal{H}_{\mathcal{M}}(\ell + 1, j) := \|\mathcal{W}\|$;
    if  $\mathcal{H}_{\mathcal{M}}(\ell + 1, j) > 0$ then
        $\ell = \ell + 1$;
        $\mathcal{V}_\ell = \mathcal{W}/\mathcal{H}_{\mathcal{M}}(\ell, j)$;
    end
    $\mathcal{W} = \mathcal{L}(\mathcal{V}_j)$;
    for $i = 1, \dots, \ell$ do
        $\mathcal{H}_{\mathcal{L}}(i, j) := \langle \mathcal{W}, \mathcal{V}_i \rangle$;
        $\mathcal{W} = \mathcal{W} - \mathcal{H}_{\mathcal{L}}(i, j)\mathcal{V}_i$;
    end
    $\mathcal{H}_{\mathcal{L}}(\ell + 1, j) := \|\mathcal{W}\|$;
    if $\mathcal{H}_{\mathcal{L}}(\ell + 1, j) > 0$ then
        $\ell = \ell + 1$;
        $\mathcal{V}_\ell = \mathcal{W}/\mathcal{H}_{\mathcal{L}}(\ell, j)$;
    end
end
```

dense and severely ill-conditioned matrices $A^{(i)}$, $i = 1, 2, 3$. In the second test example, we focus on the performances of the proposed iterative methods as a function of the distance of the matrix $\mathcal{A}$ (given by (1.6)) to the set of symmetric matrices using results from Section 3. The last example is concerned with the restoration of a hyperspectral image. For each example, we display the relative error

$$Err := \frac{\|\mathcal{X}_{\lambda_k,k} - \tilde{\mathcal{X}}\|}{\|\tilde{\mathcal{X}}\|}.$$

Here $\tilde{\mathcal{X}}$ denotes the desired solution of the error-free problem (1.3), and $\mathcal{X}_{\lambda_k,k}$ is the kth approximation computed by the algorithm used for solving (1.1) with an error-contaminated right-hand side. The regularization parameter λ_k is determined by the discrepancy principle; see below. We terminate the iterations as soon as

$$\frac{\left\|\mathcal{X}_{\lambda_k,k} - \mathcal{X}_{\lambda_{k-1},k-1}\right\|}{\left\|\mathcal{X}_{\lambda_{k-1},k-1}\right\|} \leq \tau$$

for a specified tolerance $\tau > 0$ (defined in each example), or when the maximum number of iterations $k_{\max} = 17$ is reached. The initial approximate solution in all experiments is the zero tensor. We remark that the quality of the computed solution does not change much if we decrease τ or increase $k_{\max}$; however, the number of iterations may increase. Our choices of $k_{\max}$ and τ give computed solutions of (1.1) of near-optimal quality and illustrate the relative performance of the iterative solution methods considered.

Under the table headings "Iter" and "CPU–time (sec)", we report the number of iterations and the CPU-time (in seconds) required. The regularization matrices are the tridiagonal matrices $L^{(i)} = \text{tridiag}(-1, 2, -1)$ in all examples.

We apply the discrepancy principle (with the user-chosen constant $\eta = 1.01$) to determine the regularization parameter $\lambda > 0$; see Appendix A and [24] for further details. Other methods, such as the L-curve criterion and generalized cross-validation (GCV) also may be used; see, e.g., [8, 26, 38] for discussions and references.

Example 5.1. Consider the dimensionless RTE,

$$\frac{1}{\tau_L}(\boldsymbol{\Omega}\cdot\nabla)G(\boldsymbol{r}^*,\boldsymbol{\Omega}) + G(\boldsymbol{r}^*,\boldsymbol{\Omega}) = (1-\omega)\Theta^4(\boldsymbol{r}^*) + \frac{\omega}{4\pi}\int_{4\pi} G(\boldsymbol{r}^*,\boldsymbol{\Omega}')\Phi(\boldsymbol{\Omega},\boldsymbol{\Omega}')\,\mathrm{d}\boldsymbol{\Omega}', \tag{5.1}$$

where G is the dimensionless radiative intensity. The integral in (5.1) and its boundary condition are discretized by a Nyström quadrature rule with Chebyshev–Gauss–Lobatto collocation points with the dimensionless radiative intensity approximated by an interpolation polynomial. The matrix obtained by this discretization of the dimensionless RTE can be written in the form of a Sylvester tensor equation

$$\mathcal{G} \times_1 A^{(1)} + \mathcal{G} \times_2 A^{(2)} + \mathcal{G} \times_3 A^{(3)} = \mathcal{D}, \tag{5.2}$$

where the matrices $A^{(1)}, A^{(2)}$, and $A^{(3)}$ are given in [45, Eqs. (20)-(23)]. We consider the case when $A^{(1)} = A^{(2)} = A^{(3)}$. When constructing these matrices, we choose the physically meaningful parameters $\tau_L = 1$, $\mu = \eta = \xi = 0.1$, and $m = 2$, for which the resulting coefficient matrices are nonsymmetric, dense, and highly ill-conditioned[2]. The right-hand side tensor is determined so that (5.2) has the exact solution $\tilde{\mathcal{G}} = [\tilde{g}_{ijk}]_{n\times n\times n}$, where

$$\tilde{g}_{ijk} = (x_i - \sin^2(x_i))(y_j - \sin^2(y_j))(z_k - \sin^2(z_k)),$$

where the x_i, y_j, and z_k are equidistant nodes in $[-1, 1]$ for $i, j, k = 1, 2, \ldots, n$.

The numerical results show the GAT_BTF method to give more accurate approximate solutions than the AT_BTF and FAT_BTF methods; see Table 4. GAT_BTF also requires the most iterations and, therefore, is slower than the other methods. The condition numbers reported are computed with the MATLAB command $\mathsf{cond}(\cdot)$. Notice that from Remark 2.4, we have the lower bound for the condition number

$$\text{cond}(\mathcal{A}) \geq \frac{1}{\sqrt{3}}\text{cond}(A^{(1)}),$$

which shows that the matrix $\mathcal{A}$ is extremely ill-conditioned. The values determined by the MATLAB function cond might not be very accurate, but they show that the matrices are numerically singular for all grid sizes considered.

Table 4 shows the performance of (F)AT_BTF to deteriorate as the grid size decreases. The following remark discusses the distance of matrix $\mathcal{A}$ (of the form (1.6)) to the set of symmetric matrices associated with different grid sizes.

Remark 5.2. Let $\mathcal{A}$ be the matrix with Kronecker structure (1.6) associated with the Sylvester tensor equation of Example 5.1. Table 5 reports the relative distances of $\mathcal{A}$ to the sets of all symmetric, skew-symmetric, and positive or negative semi-definite matrices.

[2] The parameter τ_L represents optical thickness, and the symbols μ, η, and ξ are, receptively, direction cosines in the x, y, and z directions. The parameter m denotes angular direction of radiation; see [45, page 092701–7].

Table 4
Results for Example 5.1 (with $\tau = 1 \cdot 10^{-2}$).

Grid	cond$(A^{(1)})$	ν	Method	Iter	Err	CPU-times(sec)
$90 \times 90 \times 90$	$4.55 \cdot 10^{16}$	0.01	AT_BTF	2	$2.48 \cdot 10^{-1}$	0.35
			FAT_BTF	2	$2.75 \cdot 10^{-1}$	0.79
			GAT_BTF	7	$6.52 \cdot 10^{-2}$	2.97
		0.001	AT_BTF	4	$1.05 \cdot 10^{-1}$	1.07
			FAT_BTF	2	$8.81 \cdot 10^{-2}$	0.85
			GAT_BTF	10	$3.62 \cdot 10^{-2}$	6.17
$120 \times 120 \times 120$	$4.86 \cdot 10^{16}$	0.01	AT_BTF	6	$4.24 \cdot 10^{-1}$	3.20
			FAT_BTF	13	$9.35 \cdot 10^{-1}$	20.11
			GAT_BTF	12	$7.72 \cdot 10^{-2}$	24.12
		0.001	AT_BTF	2	$3.82 \cdot 10^{-1}$	0.72
			FAT_BTF	7	$8.22 \cdot 10^{-1}$	6.16
			GAT_BTF	10	$5.20 \cdot 10^{-2}$	17.64
$145 \times 145 \times 145$	$1.14 \cdot 10^{17}$	0.01	AT_BTF	6	$6.23 \cdot 10^{-1}$	5.42
			FAT_BTF	5	$9.92 \cdot 10^{-1}$	6.52
			GAT_BTF	13	$4.86 \cdot 10^{-2}$	49.78
		0.001	AT_BTF	2	$5.95 \cdot 10^{-1}$	1.15
			FAT_BTF	3	$9.83 \cdot 10^{-1}$	3.41
			GAT_BTF	14	$3.97 \cdot 10^{-2}$	58.61
$200 \times 200 \times 200$	$8.21 \cdot 10^{16}$	0.01	AT_BTF	4	$8.94 \cdot 10^{-1}$	7.09
			FAT_BTF	3	$10.57 \cdot 10^{-1}$	8.89
			GAT_BTF	18	$6.03 \cdot 10^{-2}$	295.47
		0.001	AT_BTF	11	$5.90 \cdot 10^{-1}$	47.21
			FAT_BTF	5	$10.55 \cdot 10^{-1}$	16.34
			GAT_BTF	14	$6.01 \cdot 10^{-2}$	157.63

Table 5
Distances to symmetry, skew-symmetry, positive(negative) definiteness of $\mathcal{A}$ for Example 5.1.

Size($\mathfrak{D}$)	$\|\mathcal{S}(\mathcal{A})\|_2$	$\|\mathcal{H}(\mathcal{A})\|_2$	$\frac{\|\mathcal{S}(\mathcal{A})\|_2}{\|\mathcal{A}\|_{ss}}$	$\frac{\|\mathcal{H}(\mathcal{A})\|_2}{\|\mathcal{A}\|_{ss}}$	$\frac{\delta^+_{ss}(\mathcal{A})}{\|\mathcal{A}\|_{ss}}$	$\frac{\delta^-_{ss}(\mathcal{A})}{\|\mathcal{A}\|_{ss}}$
$90 \times 90 \times 90$	65.1847	93.7889	0.4100	0.5900	0.9874	1.0000
$120 \times 120 \times 120$	116.5079	166.8792	0.4111	0.5889	0.9929	1.0000
$145 \times 145 \times 145$	170.5865	243.8935	0.4116	0.5884	0.9952	1.0000
$200 \times 200 \times 200$	325.7488	464.8630	0.4120	0.5880	0.9975	1.0000

Recall that $\|\mathcal{A}\|_{ss} = \|\mathcal{S}(\mathcal{A})\|_2 + \|\mathcal{H}(\mathcal{A})\|_2$, where $\mathcal{H}(\mathcal{A})$ and $\mathcal{S}(\mathcal{A})$ denote the symmetric and skew-symmetric parts of $\mathcal{A}$, respectively. The table shows the relative distances of the matrix $\mathcal{A}$ to both the sets of symmetric and skew-symmetric matrices to converge to 0.5 as the mesh size decreases. This may be a reason for the poor performance of the (F)AT_BTF methods in the previous example.

The observation of Remark 5.2 motivated us to further examine the dependence of the performances of the proposed algorithms to the distance of $\mathcal{A}$ to the set of symmetric matrices. We note that in the previous test problem, equation (1.1) has a solution of low rank. However, the right-hand side is not a tensor of low-rank since it is contaminated by error. In the following test problem, we examine the performances of (F)AT_BTF for the cases when the exact solution is of low rank and when it is not. The coefficient matrices $A^{(i)}$ are sparse.

Example 5.3. *Consider the Sylvester tensor equation* (1.1), *in which* $A^{(l)} = (I+\alpha S)H_l$ *for* $l = 1, 2, 3$. *We let* $S = \text{tridiag}(-1, 0, 1) \in \mathbb{R}^{n \times n}$. *The matrices* $H_l = [h^{(l)}_{ij}] \in \mathbb{R}^{n \times n}$ *are*

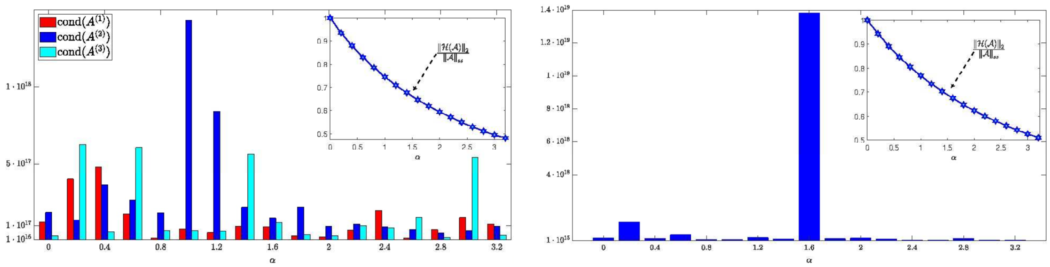

FIG. 2. *Outside figure: "cond($A^{(i)}$)" (left) and "lower bound for cond($\mathcal{A}$)" (right) versus α; Inside figure: Relative distance of $\mathcal{A} \in \mathbb{R}^{n^3 \times n^3}$ to the symmetric matrices versus α for $n = 450$ (left) and $n = 100$ (right).*

Toeplitz matrices given by

$$h_{ij}^{(l)} = \begin{cases} \dfrac{1}{2r-1}, & |i-j| \le r, \\ 0, & \textit{otherwise}, \end{cases} \tag{5.3}$$

for $l = 1, 2, 3$. Note that $\mathcal{A} \in \mathbb{R}^{n^3 \times n^3}$. We consider two values of n:

- *When $n = 100$, we set $r = 6$ for H_1, H_2, and H_3. The noise level was $\nu = 0.01$ and we set $\tau = 3 \cdot 10^{-2}$. We were able to compute the lower bound for* cond*($\mathcal{A}$) by using Remark 2.4. This bound is reported in Figure 2.*
- *When $n = 450$, we set $r = 5$ for H_1, $r = 6$ for H_2, and $r = 5.5$ for H_3. The noise level was $\nu = 0.01$ and we set $\tau = 6 \cdot 10^{-2}$.*

For each dimension, Figure 2 depicts the relative distance of $\mathcal{A}$ to the set of symmetric matrices together with information about condition numbers. Note that the distance depends on the parameter α. Equations (1.1) with the two exact solutions of the associated error-free problem are considered:

Case I: *Let $\tilde{x} = \tilde{x}_1 \otimes \tilde{x}_2 \otimes \tilde{x}_3$, where $\tilde{x}_i = \text{rand}(n, 1)$ for $i = 1, 2, 3$. The MATLAB function* rand *generates uniformly distributed random numbers in the interval $[0, 1]$. We determine the error-free right-hand side so that $\tilde{\mathcal{X}}$, with $\text{vec}(\tilde{\mathcal{X}}) = \tilde{x}_1 \otimes \tilde{x}_2 \otimes \tilde{x}_3$, is the exact solution of (1.1). Thus, the solution of the error-free equation associated with eq. (1.1) has low rank.*

Case II: *We determine the error-free right-hand side so that $\tilde{\mathcal{X}} = \text{rand}(n, n, n)$ is the solution of the error-free equation associated with (1.1).*

Figures 3 and 4 show the performance of the iterative methods for different values of α. The quality of the computed approximate solutions are shown in the right-hand side plots and the required CPU-times in the left-hand side plots. For Case I, where the problem has a low-rank solution, the relative performances of the methods agree with the observations of Remark 5.2, i.e., poor performance of (F)AT_BTF is observed as the relative distance of $\mathcal{A}$ tends to 0.5. The situation for matrices of Class II is more complicated. This illustrates that the performance of (F)AT_BTF does depend on the properties of the right-hand side and, therefore, on the properties of the desired solution.

Differently from Example 5.1, the AT_BTF and FAT_BTF methods determine approximate solutions of, generally, higher quality than the GAT_BTF method for test Examples 5.3 and 5.4.

Example 5.4. The exact solution of this test example is a $1017 \times 1340 \times 33$ tensor that represents a hyperspectral image; see [17, 30, 43] for discussions on hyperspectral image restoration. Each slice of the solution tensor corresponds to an image of the same scene measured at a different wavelength. We consider the following Sylvester tensor equation

$$\mathcal{X} \times_1 A^{(1)} + \mathcal{X} \times_2 A^{(2)} + \mathcal{X} \times_3 A^{(3)} = \mathcal{D},$$

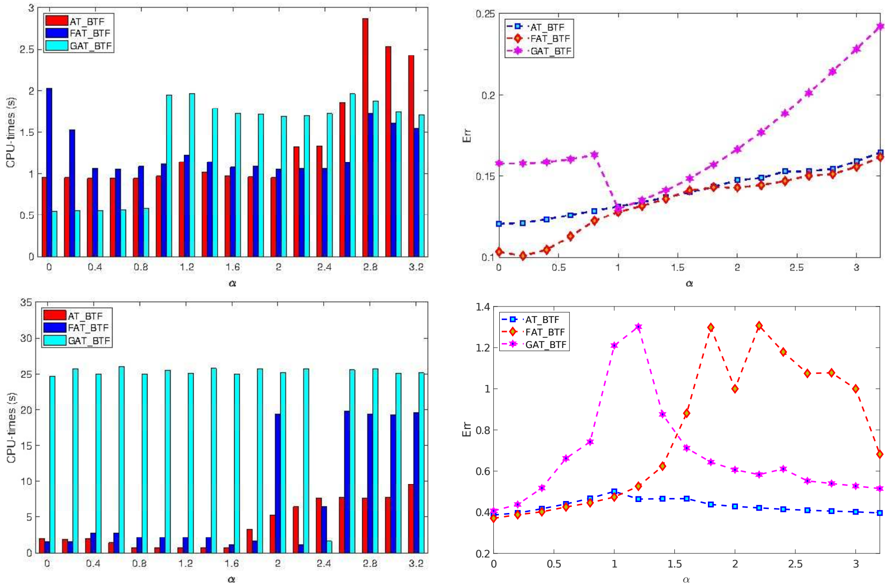

Fig. 3. *Performance of the iterative methods applied to the matrices of Example 5.3; Cases I (top) and II (bottom) for $n = 100$.*

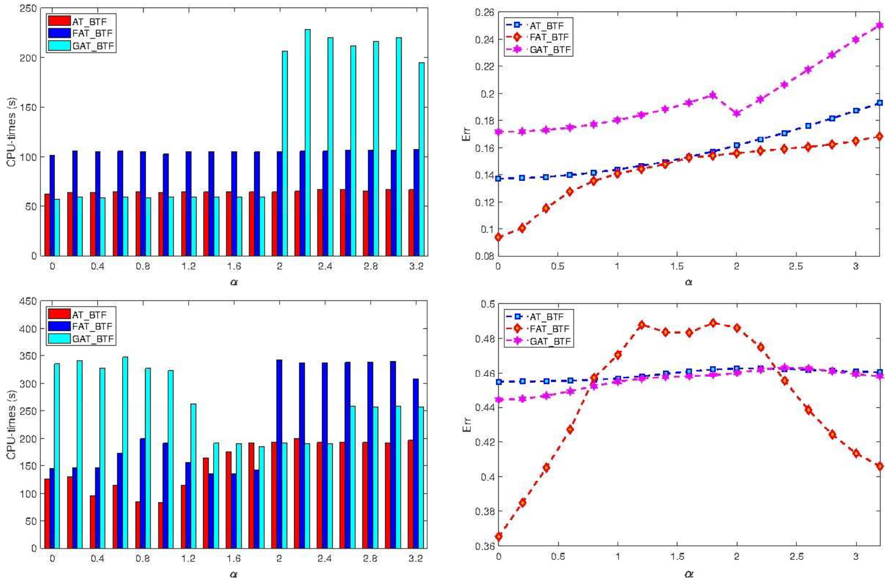

Fig. 4. *Performance of the iterative methods applied to the matrices of Example 5.3; Cases I (top) and II (bottom) for $n = 450$.*

where $A^{(1)} = [a_{ij}^{(1)}]$, $A^{(2)} = [a_{ij}^{(2)}]$, and $A^{(3)} = [a_{ij}^{(3)}]$ are 1017×1017, 1340×1340, and 33×33 matrices, respectively, in the form $A^{(l)} = (I + \alpha R_l)H_l$ with H_l given by (5.3), and R_l has uniformly distributed random entries in the interval $[0, 1]$. The dimensions are suitably chosen for $l = 1, 2, 3$. We set $r = 5$ for $A^{(1)}$, and $r = 7$ for $A^{(2)}$ and $A^{(3)}$. The condition numbers for these matrices are reported in Table 6. When $\alpha = 0$, the coefficient matrices $A^{(l)}$ reduce to blurring matrices exploited in [8].

In Table 6, LT_BTF stands for the method obtained by replacing the Arnoldi process

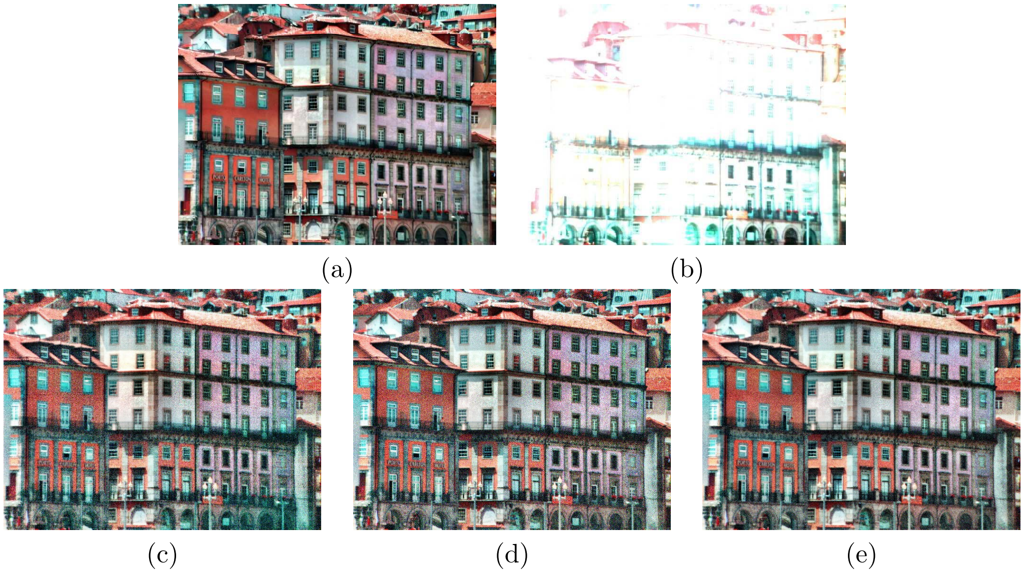

Fig. 5. *(a) Exact solution, (b) Noisy data, (c) Restored data with GAT_BTF, (d) FAT_BTF and (e) AT_BTF for level of noise $\nu = 0.01$ and $\alpha = 0.001$.*

in AT_BTF by the Lanczos process, which requires less arithmetic work. This replacement is possible because the matrices $A^{(l)}$ are symmetric for $\alpha = 0$. We remark that also the GAT_BTF method can be simplified when all matrices are symmetric. Since the GAT_BTF method gives restorations of worse quality than LT_BTF, we will not dwell on this simplification. Figure 5 shows the uncontaminated image that we would like to determine, as well as the contaminated image that is assumed to be available. Restorations achieved by the AT_BTF, FAT_BTF, and GAT_BTF methods also are displayed in Figure 5.

We report the distances of the matrices $\mathcal{A}$ to the sets of symmetric, skew-symmetric, and positive (negative) definite matrices in Table 7. It can be seen that for small $\alpha > 0$, the matrix $\mathcal{A}$ is almost symmetric. For $\alpha = 0.001$, the coefficient matrices are dense and the AT_BTF and GAT_BTF methods perform similarly. The FAT_BTF method produces more accurate solutions, but requires more CPU-time than the other methods. The matrices $A^{(l)}$ are sparse and symmetric when $\alpha = 0$. Then the LT_BTF can be applied. This method is faster than the other methods.

Table 6
Results for Example 5.4 (with $\tau = 3 \cdot 10^{-2}$).

α	cond$(A^{(i)})$	Method	Iter	Err	CPU-times(sec)
0	cond$(A^{(1)}) = 2.32 \cdot 10^{17}$	LT_BTF	2	$4.42 \cdot 10^{-2}$	15.07
	cond$(A^{(2)}) = 1.48 \cdot 10^{18}$	FAT_BTF	2	$3.49 \cdot 10^{-2}$	50.82
	cond$(A^{(3)}) = 2.96 \cdot 10^{17}$	GAT_BTF	2	$4.45 \cdot 10^{-2}$	27.09
0.001	cond$(A^{(1)}) = 1.52 \cdot 10^{17}$	AT_BTF	3	$5.71 \cdot 10^{-2}$	30.79
	cond$(A^{(2)}) = 9.33 \cdot 10^{17}$	FAT_BTF	2	$3.84 \cdot 10^{-2}$	52.03
	cond$(A^{(3)}) = 1.58 \cdot 10^{17}$	GAT_BTF	2	$5.70 \cdot 10^{-2}$	28.67

Table 7

Distances to symmetry, skew-symmetry, positive (negative) definiteness of $\mathcal{A}$ for Example 5.4.

α	$\|\mathcal{S}(\mathcal{A})\|_2$	$\|\mathcal{H}(\mathcal{A})\|_2$	$\frac{\|\mathcal{S}(\mathcal{A})\|_2}{\|\mathcal{A}\|_{ss}}$	$\frac{\|\mathcal{H}(\mathcal{A})\|_2}{\|\mathcal{A}\|_{ss}}$	$\frac{\delta^+_{ss}(\mathcal{A})}{\|\mathcal{A}\|_{ss}}$	$\frac{\delta^-_{ss}(\mathcal{A})}{\|\mathcal{A}\|_{ss}}$
0	0	3.4558	0	1	0.2143	1.0000
0.001	0.0272	4.8596	0.0056	0.9944	0.1572	1.0000

6. Conclusions.

This paper considers linear systems of equations with a matrix with the structure

$$\mathcal{A} = \sum_{j=1}^{N} I^{(I_N)} \otimes \ldots \otimes I^{(I_{j+1})} \otimes A^{(j)} \otimes I^{(I_{j-1})} \otimes \ldots \otimes I^{(I_1)}. \tag{6.1}$$

An extension is discussed in Remark 2.9. We first show some bounds for the condition number of the matrix (6.1) and discuss ways of measuring the distance of this matrix to the sets of (skew-)symmetric and positive (negative) definite matrices. These results are then used to compare several iterative solution methods for very ill-conditioned Sylvester tensor equations. These methods are based on the Arnoldi process, the flexible Arnoldi process, or a generalized Arnoldi process. Tikhonov regularization is applied. The iterative methods considered generalize methods discussed in [8, 24, 38], as well as the flexible Arnoldi process [41]. Numerical examples with applications to the solution of a radiative transfer equation in 3D and to color image restoration illustrate that approximate solutions of high quality can be computed with fairly few iteration steps and, hence, with fairly little computational effort.

Acknowledgments. Research by LR was supported in part by NSF grants DMS-1720259 and DMS-1729509. We would like to thank the referees for their useful comments.

REFERENCES

[1] M. August, M. C. Bañuls, and T. Huckle, On the approximation of functionals of very large Hermitian matrices represented as matrix product operators, Electronic Transactions on Numerical Analysis, 46 (2017), pp. 215–232.

[2] B. W. Bader and T. G. Kolda, MATLAB Tensor Toolbox Version 2.5. `http://www.sandia.gov/~tgkolda/TensorToolbox`.

[3] Z. Z. Bai, G. H. Golub, and M. K. Ng, Hermitian and skew-Hermitian splitting methods for non-Hermitian positive definite linear systems, SIAM Journal on Matrix Analysis and Applications, 24 (2003), pp. 603–626.

[4] J. Ballani and L. Grasedyck, A projection method to solve linear systems in tensor format, Numerical Linear Algebra with Applications, 20 (2013), pp. 27–43.

[5] F. P. A. Beik, F. S. Movahed, and S. Ahmadi-Asl, On the Krylov subspace methods based on tensor format for positive definite Sylvester tensor equations, Numerical Linear Algebra with Applications, 23 (2016), pp. 444–466.

[6] A. H. Bentbib, M. El Guide, K. Jbilou, E. Onunwor, and L. Reichel, Solution methods for linear discrete ill-posed problems for color image restoration, BIT Numerical Mathematics, 58 (2018), pp. 555–578.

[7] J. Blanco, O. Rojas, C. Chacón, J. M. Guevara-Jordan, and J. Castillo, Tensor formulation of 3-D mimetic finite differences and applications to elliptic problems, Electronic Transactions on Numerical Analysis, 45 (2016), pp. 457–475.

[8] A. Bouhamidi, K. Jbilou, L. Reichel, and H. Sadok, A generalized global Arnoldi method for ill-posed matrix equations, Journal of Computational and Applied Mathematics, 236 (2012), pp. 2078–2089.

[9] A. Buccini, M. Pasha, and L. Reichel, Generalized singular value decomposition with iterated Tikhonov regularization, Journal of Computational and Applied Mathematics, in press.

[10] D. Calvetti, B. Lewis, and L. Reichel, On the regularizing properties of the GMRES method, Numerische Mathematik, 91 (2002), pp. 605–625.

[11] Z. Chen and L. Z. Lu, A projection method and Kronecker product preconditioner for solving Sylvester tensor equations, Science China Mathematics, 55 (2012), pp. 1281–1292.

[12] A. Cichocki, R. Zdunek, A. H. Phan, and S. I. Amari, Nonnegative Matrix and Tensor Factorizations: Applications to Exploratory Multi-Way Data Analysis and Blind Source Separation, Wiley, Chichester, 2009.

[13] M. Donatelli, D. Martin, and L. Reichel, Arnoldi methods for image deblurring and anti-reflective boundary conditions, Applied Mathematics and Computation, 253 (2015), pp. 135–150.

[14] K. Du, J. Duintjer Tebbens, and G. Meurant, Any admissible harmonic Ritz value set is possible for GMRES, Electronic Transactions on Numerical Analysis, 47 (2017), pp. 37–56.

[15] L. Dykes, S. Noschese, and L. Reichel, Rescaling the GSVD with application to ill-posed problems, Numerical Algorithms, 68 (2015), pp. 531–545.

[16] H. Y. Fan, L. Zhang, E. K. Chu, and Y. Wei, Numerical solution to a linear equation with tensor product structure, Numerical Linear Algebra with Applications, 24 (2017), e2106.

[17] D. H. Foster, K. Amano, S. M. C. Nascimento, and M. J. Foster, Frequency of metamerism in natural scenes, Journal of the Optical Society of America A, 23 (2006), pp. 2359–2372.

[18] S. Gazzola, S. Noschese, P. Novati, and L. Reichel, Arnoldi decomposition, GMRES, and preconditioning for linear discrete ill-posed problems, Applied Numerical Mathematics, 142 (2019) pp. 102–121.

[19] S. Gazzola and J. G. Nagy, Generalized Arnoldi–Tikhonov method for sparse reconstruction, SIAM J. Sci. Comput., 36 (2014), pp. B225–B247.

[20] P. C. Hansen, Rank-Deficient and Discrete Ill-Posed Problems, SIAM, Philadelphia, 1998.

[21] N. J. Higham, Computing a nearest symmetric positive semidefinite matrix, Linear Algebra and its Applications, 103, (1988), pp. 103–118.

[22] R. B. Holmes, Best approximation by normal operators, Journal of Approximation Theory, 12 (1974), pp. 412–417.

[23] R. A. Horn and C. R. Johnson, Matrix Analysis, Cambridge University Press, Cambridge, 1985.

[24] G. Huang, L. Reichel, and F. Yin, On the choice of subspace for large-scale Tikhonov regularization problems in general form, Numerical Algorithms, 81 (2019), pp. 33–55.

[25] B. N. Khoromskij, Tensors-structured numerical methods in scientific computing: Survey on recent advances, Chemometrics and Intelligent Laboratory Systems, 110 (2012), pp. 1–19.

[26] S. Kindermann, Convergence analysis of minimization-based noise level-freeparameter choice rules for linear ill-posed problems, Electronic Transactions on Numerical Analysis, 38 (2011), pp. 233–257.

[27] T. G. Kolda and B. W. Bader, Tensor decompositions and applications, SIAM Review, 51 (2009), pp. 455–500.

[28] D. Kressner and C. Tobler, Low-rank tensor Krylov subspace methods for parametrized linear systems, SIAM Journal on Matrix Analysis and Applications, 32 (2011), pp. 1288–1316.

[29] B. W. Li, S. Tian, Y. S. Sun, and Z. M. Hu, Schur-decomposition for $3D$ matrix equations and its application in solving radiative discrete ordinates equations discretized by Chebyshev collocation spectral method, Journal of Computational Physics, 229 (2010), pp. 1198–1212.

[30] F. Li, M. K. Ng, and R. J. Plemmons, Coupled segmentation and denoising/deblurring for hyperspectral material identification, Numerical Linear Algebra with Applications, 19 (2012), pp. 153–173.

[31] N. Li, C. Navasca, and C. Glenn, Iterative methods for symmetric outer product tensor decomposition, Electronic Transactions on Numerical Analysis, 44 (2015), pp. 124–139.

[32] R.-C. Li and Q. Ye, A Krylov subspace method for quadratic matrix polynomials with application to constrained least squares problems, SIAM Journal on Matrix Analysis and Applications, 25 (2003), pp. 405–428.

[33] L. Liang and B. Zheng, Sensitivity analysis of the Lyapunov tensor equation, Linear and Multilinear Algebra, 67 (2019), pp. 555–572.

[34] A. Malek, Z. K. Bojdi, and P. N. N. Golbarg, Solving fully three-dimensional microscale dual phase lag problem using mixed-collocation finite difference discretization, Journal of Heat Transfer, 134 (2012), 094504.

[35] A. Malek and S. H. M. Masuleh, Mixed collocation-finite difference method for 3D microscopic heat transport problems, Journal of Computational and Applied Mathematics, 217 (2008), pp. 137–147.

[36] S. H. M. Masuleh and T. N. Phillips, Viscoelastic flow in an undulating tube using spectral methods, Computers & Fluids, 33 (2004), pp. 1075–1095.

[37] K. Morikuni, L. Reichel, and K. Hayami, FGMRES for linear discrete ill-posed problems, Applied Numerical Mathematics, 75 (2014), pp. 175–187.

[38] L. Reichel, F. Sgallari, and Q. Ye, Tikhonov regularization based on generalized Krylov subspace methods, Applied Numerical Mathematics, 62 (2012), pp. 1215–1228.

[39] L. Reichel and A. Shyshkov, A new zero-finder for Tikhonov regularization, BIT Numerical Mathematics, 48 (2008), pp. 627-643.

[40] L. Reichel and X. Yu, Tikhonov regularization via flexible Arnoldi reduction, BIT Numerical Mathematics, 55 (2015), pp. 1145–1168.

[41] Y. Saad, Iterative Methods for Sparse Linear Systems, 2nd ed., SIAM, 2003.

[42] X. Shi, Y. Wei, and S. Ling, Backward error and perturbation bounds for high order Sylvester tensor equation, Linear and Multilinear Algebra, 61 (2013), pp. 1436–1446.

[43] M. Signoretto, R. Van de Plas, B. De Moor, and J. A. K. Suykens, Tensor versus matrix completion:

A comparison with application to spectral data, IEEE Signal Processing Letters, 18 (2011), pp. 403–406.

[44] V. Simoncini, Computational methods for linear matrix equations, SIAM Review, 58 (2016), pp. 377–441.

[45] Y. S. Sun, M. Jing, and B. W. Li, Chebyshev collocation spectral method for three-dimensional transient coupled radiative-conductive heat transfer, Journal of Heat Transfer, 134 (2012), pp. 092701–092707.

[46] M. K. Zak and F. Toutounian, Nested splitting conjugate gradient method for matrix equation $AXB = C$ and preconditioning, Computers & Mathematics with Applications, 66 (2013), pp. 269–278.

Appendix A. We briefly describe how the discrepancy principle can be used for determining a suitable value of the regularization parameter in the proposed algorithms. We will use the notation of Algorithm 2. The algorithms reduce the minimization problem (1.4) to a low-dimensional problem of the form

$$\min_{z\in\mathbb{R}^k} \left\{ \left\| \|\mathcal{D}\| e_1 - \hat{H}_k z \right\|_2^2 + \lambda \left\| z \right\|_2^2 \right\} \tag{6.2}$$

for some $\lambda > 0$. The tensor $\mathcal{D}$ is contaminated by an error $\mathcal{E}$; cf. (1.2). Assume that a bound $\delta > 0$ for the norm of $\mathcal{E}$ is available, i.e.,

$$\|\mathcal{E}\| \leq \delta.$$

Then the discrepancy principle can be used to determine the regularization parameter λ. The discrepancy principle prescribes that $\lambda > 0$ is chosen so that the solution $z_{\lambda,k}$ of (6.2) satisfies

$$\left\| \|\mathcal{D}\| e_1 - \hat{H}_k z_{\lambda,k} \right\|_2^2 = \eta^2 \delta^2, \tag{6.3}$$

where $\eta > 1$ is a user-chosen parameter that is independent of δ. The vector $z_{\lambda,k}$ is known in closed form. Substituting this vector into (6.3) and using the singular value decomposition of $\hat{H}_k$ gives a simple equation for λ; see, e.g., [24] for related formulas. This equation can be solved quite inexpensively by, e.g., Newton's method. Other zero-finders are descibed in [9, 39].