Electronic Transactions on Numerical Analysis.
Volume 42, pp. 147-164, 2014.

ISSN 1068-9613.

ETNA
Kent State University
http://etna.math.kent.edu

RATIONAL INTERPOLATION METHODS FOR SYMMETRIC SYLVESTER EQUATIONS*

PETER BENNER† AND TOBIAS BREITEN‡

Dedicated to Lothar Reichel on the occasion of his 60th birthday

Abstract. We discuss low-rank approximation methods for large-scale symmetric Sylvester equations. Following similar discussions for the Lyapunov case, we introduce an energy norm by the symmetric Sylvester operator. Given a rank n_r, we derive necessary conditions for an approximation being optimal with respect to this norm. We show that the norm minimization problem is related to an objective function based on the $\mathcal{H}_2$-inner product for symmetric state space systems. This objective function leads to first-order optimality conditions that are equivalent to the ones for the norm minimization problem. We further propose an iterative procedure and demonstrate its efficiency by means of some numerical examples.

Key words. Sylvester equations, rational interpolation, energy norm

AMS subject classifications. 15A24, 37M99

1. Introduction.

In this paper, we consider large-scale linear matrix equations

$$\mathbf{AXM} + \mathbf{EXH} = \mathbf{G}, \tag{1.1}$$

where $\mathbf{A}, \mathbf{E} \in \mathbb{R}^{n\times n}, \mathbf{M}, \mathbf{H} \in \mathbb{R}^{m\times m}$, and $\mathbf{G} \in \mathbb{R}^{n\times m}$. The sought-after solution $\mathbf{X} \in \mathbb{R}^{n\times m}$ to the *Sylvester equation* (1.1) is of great interest within systems and control theory; see [1]. In particular, for $\mathbf{M} = \mathbf{E}^T$, $\mathbf{H} = \mathbf{A}^T$, and $\mathbf{G} = \mathbf{BB}^T$, the resulting *Lyapunov equation* characterizes stability properties of an underlying dynamical system

$$\begin{aligned} \mathbf{E}\dot{\mathbf{x}}(t) &= \mathbf{Ax}(t) + \mathbf{Bu}(t), \\ \mathbf{y}(t) &= \mathbf{Cx}(t), \end{aligned} \tag{1.2}$$

where, respectively, $\mathbf{x}(t), \mathbf{u}(t)$, and $\mathbf{y}(t)$ are called *state*, *control*, and *output* of the system. Linear matrix equations of the form (1.1) have been studied for several years now. However, finding efficient algorithms for large n, m is still an active area of research within the numerical linear algebra community. For a detailed introduction into linear matrix equations, we refer to the two recent survey articles [13, 39]. Since direct methods, e.g., the Bartels-Stewart algorithm [5] or Hammarling's method [28] require cubic complexity to solve (1.1), they are only feasible as long as n, m are of medium size. Depending on the individual computer architecture, this nowadays might cover system dimensions up to $n, m \sim 10^4$. Often, however, dynamical systems and thus matrix equations result from a spatial discretization of a partial differential equation (PDE). Here, one easily ends up with dimensions that cannot be handled by the mentioned direct methods. For the general case where $\mathbf{G}$ is of full rank, there is still no easily applicable technique to compute $\mathbf{X}$. On the other hand, assuming that $\mathbf{G} = \mathbf{BC}^T$, where $\operatorname{rank}(\mathbf{B}), \operatorname{rank}(\mathbf{C}) \ll n, m$, the singular values of $\mathbf{X}$ often decay very fast; see [3, 25, 32, 36]. In other words, the low numerical rank of the solution allows for *low-rank approximations* $\mathbf{X} \approx \mathbf{V}\mathbf{X}_r\mathbf{W}^T$, where $\mathbf{V} \in \mathbb{R}^{n\times n_r}, \mathbf{W} \in \mathbb{R}^{m\times n_r}$,

*Received November 21, 2013. Accepted July 22, 2014. Published online on September 29, 2014. Recommended by Valeria Simoncini. Most of this work was completed while the second author was at the Max Planck Institute for Dynamics of Complex Technical Systems, Magdeburg.

†Computational Methods in Systems and Control Theory, Max Planck Institute for Dynamics of Complex Technical Systems, Sandtorstr. 1, 39106 Magdeburg, Germany (benner@mpi-magdeburg.mpg.de).

‡Institute for Mathematics and Scientific Computing, Heinrichstr. 36/III, University of Graz, Austria (tobias.breiten@uni-graz.at).

ETNA
Kent State University
http://etna.math.kent.edu

and $\mathbf{X}_r \in \mathbb{R}^{n_r \times n_r}$, with $n_r \ll n, m$. This phenomenon has lead to several numerically efficient methods that are also applicable in a large-scale setting. The most popular choices can basically be classified into two categories: (a) methods based on alternating directions implicit (ADI) schemes; (b) methods based on projection and prolongation. Methods that specifically address the computation of low-rank approximations to the solution of Sylvester equations can be found in [6, 8, 11, 21]. The literature on low-rank solution for the Lyapunov case goes back even further and has achieved more attention; for a detailed overview on this topic, we refer to [10, 12, 13, 19, 30, 31, 33, 35, 37, 38, 42]. Other techniques are based on the tensorized linear system (see [14, 24, 32]) or Riemannian optimization; see [40, 41]. Especially the latter class of methods is important in the context of this article as it inspired the approach discussed here though we do not use Riemannian optimization explicitly. It rather occurs implicitly at the minimization of a certain energy norm.

The structure of this paper is as follows. For a special symmetry property of the matrices in (1.1), in Section 2 we introduce an objective function based on the energy norm of the underlying Sylvester operator. We further derive first-order necessary conditions for this objective function. In Section 3, we establish a connection between the energy norm and the $\mathcal{H}_2$-inner product of two dynamical control systems of the form (1.2). We show that this inner product exhibits first-order necessary optimality conditions that are equivalent to the ones for the energy norm. Based on techniques from rational interpolation, we discuss the use of an iterative Sylvester solver applicable in large-scale settings. In Section 4, we provide numerical results to demonstrate the applicability of the method. As these results correspond to Sylvester equations arising in imaging, we briefly review the use of large-scale Sylvester equations (1.1) for problems evolving in image restoration as discussed in [15, 18]. We conclude with a short summary in Section 5.

In all what follows, $\mathbf{A} \succ 0$ ($\mathbf{A} \succeq 0$) denotes a symmetric positive (semi-)definite matrix. With $\otimes$ we denote the Kronecker product of two matrices. Vectorization of a matrix $\mathbf{A}$, i.e., stacking all columns of $\mathbf{A}$ into a long vector, is denoted by $\operatorname{vec}(\mathbf{A})$. The (matrix-valued) residue of a meromorphic matrix-valued function $\mathbf{G}(s)$ at a point $\lambda \in \mathbb{C}$ is denoted as $\operatorname{res}[\mathbf{G}(s), \lambda]$. All vectors and matrices are denoted by boldface letters and scalar quantities by italic letters. The Kronecker delta δ_{ij} is defined as

$$\delta_{ij} := \begin{cases} 1 & i = j, \\ 0 & \text{otherwise.} \end{cases}$$

2. Symmetric Sylvester equations and the energy norm. From now on, we consider symmetric Sylvester equations of the form

$$\mathbf{AXM} + \mathbf{EXH} = \mathbf{G}, \tag{2.1}$$

where $\mathbf{A}, \mathbf{E} \in \mathbb{R}^{n \times n}, \mathbf{A}, \mathbf{E} \succ 0$, $\mathbf{M}, \mathbf{H} \in \mathbb{R}^{m \times m}, \mathbf{M}, \mathbf{H} \succ 0$, and $\mathbf{G} \in \mathbb{R}^{n \times m}$. While in some applications the matrix $\mathbf{G}$ is not necessarily of low (numerical) rank, we still might construct approximations $\mathbf{X} \approx \tilde{\mathbf{X}} := \mathbf{V}\mathbf{X}_r\mathbf{W}^T$ with $\mathbf{V} \in \mathbb{R}^{n \times n_r}, \mathbf{W} \in \mathbb{R}^{m \times n_r}$, and $\mathbf{X}_r \in \mathbb{R}^{n_r \times n_r}$. Note that we do not require $\mathbf{X}_r$ to be a square matrix and thus we have the freedom to choose $\mathbf{V}$ and $\mathbf{W}$ such that they have a different number of columns. Still, using $\mathbf{X}_r \in \mathbb{R}^{n_r \times n_r}$ seems to be a natural choice and also simplifies the notation. This representation can always be obtained from a rectangular $\mathbf{X}_r$ by employing its singular value decomposition (SVD). Throughout the article, we always assume that $\mathbf{V}$ and $\mathbf{W}$ have full column rank and $\mathbf{X}_r$ is nonsingular.

The most common way to evaluate the quality of an approximation is by means of the norm of the error $\|\mathbf{X} - \tilde{\mathbf{X}}\|$. For the spectral norm or the Frobenius norm, the best rank n_r

ETNA
Kent State University
http://etna.math.kent.edu

approximation is given by the SVD. This result is well-known and follows from the Eckart-Young-Mirsky theorem that can be found in standard textbooks such as, e.g., [23]. Unfortunately, computing an SVD-based approximation would require the full solution $\mathbf{X}$ itself. For symmetric systems, however, another natural choice for measuring errors is the energy norm. Note that due to the definiteness of the matrices, for the error $\mathbf{X} - \tilde{\mathbf{X}}$ we can define a norm via

$$\|\mathbf{X} - \tilde{\mathbf{X}}\|_{\mathcal{L}_S}^2 := \underbrace{\operatorname{vec}\left(\mathbf{X} - \tilde{\mathbf{X}}\right)^T}_{\mathbf{e}^T} \underbrace{(\mathbf{M} \otimes \mathbf{A} + \mathbf{H} \otimes \mathbf{E})}_{=:\mathcal{L}_S \succ 0} \underbrace{\operatorname{vec}\left(\mathbf{X} - \tilde{\mathbf{X}}\right)}_{\mathbf{e}}.$$

The energy norm for matrix equations was first investigated in detail in [40, 41] and later discussed in the context of $\mathcal{H}_2$-model reduction in [9]. Note that there is also a direct connection between the Frobenius norm and the energy norm of the error $\mathbf{X} - \tilde{\mathbf{X}}$:

$$\|\mathbf{X} - \tilde{\mathbf{X}}\|_{\mathcal{L}_S}^2 = \mathbf{e}^T \mathcal{L}_S \mathbf{e} = \frac{\mathbf{e}^T \mathcal{L}_S \mathbf{e}}{\mathbf{e}^T \mathbf{e}} \|\mathbf{e}\|_2^2 \geq \lambda_{\min}(\mathcal{L}_S) \, \|\mathbf{X} - \tilde{\mathbf{X}}\|_F^2.$$

The previous inequality holds due to the fact that the Rayleigh quotient $R(\mathcal{L}_S, \mathbf{e})$ is bounded from below by the minimal eigenvalue of the symmetric matrix $\mathcal{L}_S$. Assume now that for a given Sylvester equation (2.1) and a prescribed dimension $n_r \ll n$, the goal is to find matrices $\mathbf{V} \in \mathbb{R}^{n \times n_r}$, $\mathbf{W} \in \mathbb{R}^{m \times n_r}$, and $\mathbf{X}_r \in \mathbb{R}^{n_r \times n_r}$ such that

$$\|\mathbf{X} - \mathbf{V}\mathbf{X}_r\mathbf{W}^T\|_{\mathcal{L}_S}^2 = \min_{\substack{\tilde{\mathbf{V}} \in \mathbb{R}^{n \times n_r},\, \tilde{\mathbf{W}} \in \mathbb{R}^{m \times n_r}, \\ \tilde{\mathbf{X}}_r \in \mathbb{R}^{n_r \times n_r} \text{ nonsingular}}} \|\mathbf{X} - \tilde{\mathbf{V}}\tilde{\mathbf{X}}_r\tilde{\mathbf{W}}^T\|_{\mathcal{L}_S}^2. \tag{2.2}$$

As a first step towards optimization, one usually considers first-order necessary optimality conditions for $\mathbf{V}, \mathbf{W}$, and $\mathbf{X}_r$. For this, we state some useful properties for computing the derivative of the trace function with respect to a matrix. According to [4], for a matrix $\mathbf{Y} \in \mathbb{R}^{n \times m}$ and matrices $\mathbf{K}, \mathbf{L}$ of compatible dimensions, it holds that

$$\begin{aligned} \frac{\partial}{\partial \mathbf{Y}} \left[\operatorname{tr}(\mathbf{K}\mathbf{Y}\mathbf{L})\right] &= \mathbf{K}^T\mathbf{L}^T, \\ \frac{\partial}{\partial \mathbf{Y}} \left[\operatorname{tr}\left(\mathbf{K}\mathbf{Y}\mathbf{L}\mathbf{Y}^T\right)\right] &= \mathbf{K}^T\mathbf{Y}\mathbf{L}^T + \mathbf{K}\mathbf{Y}\mathbf{L}. \end{aligned} \tag{2.3}$$

Using these properties, we can give the following generalization of results similarly obtained for the Lyapunov equation in [41].

LEMMA 2.1. *Assume that* $(\mathbf{V}, \mathbf{W}, \mathbf{X}_r)$ *solves* (2.2)*. Then it holds*

$$\left(\mathbf{A}\mathbf{V}\mathbf{X}_r\mathbf{W}^T\mathbf{M} + \mathbf{E}\mathbf{V}\mathbf{X}_r\mathbf{W}^T\mathbf{H} - \mathbf{G}\right)\mathbf{W} = \mathbf{0}, \tag{2.4a}$$

$$\mathbf{V}^T\left(\mathbf{A}\mathbf{V}\mathbf{X}_r\mathbf{W}^T\mathbf{M} + \mathbf{E}\mathbf{V}\mathbf{X}_r\mathbf{W}^T\mathbf{H} - \mathbf{G}\right) = \mathbf{0}, \tag{2.4b}$$

$$\mathbf{V}^T\left(\mathbf{A}\mathbf{V}\mathbf{X}_r\mathbf{W}^T\mathbf{M} + \mathbf{E}\mathbf{V}\mathbf{X}_r\mathbf{W}^T\mathbf{H} - \mathbf{G}\right)\mathbf{W} = \mathbf{0}. \tag{2.4c}$$

Proof. Note that by vectorization of (2.1), we know that

$$\mathcal{L}_S \operatorname{vec}(\mathbf{X}) = \operatorname{vec}(\mathbf{G}).$$

ETNA
Kent State University
http://etna.math.kent.edu

Consequently, we obtain

$$\begin{aligned} f(\mathbf{V},\mathbf{W},\mathbf{X}_r) &= \operatorname{vec}\left(\mathbf{X}-\mathbf{V}\mathbf{X}_r\mathbf{W}^T\right)^T \mathcal{L}_S \operatorname{vec}\left(\mathbf{X}-\mathbf{V}\mathbf{X}_r\mathbf{W}^T\right) \\ &= \operatorname{vec}(\mathbf{X})^T \operatorname{vec}(\mathbf{G}) - 2\operatorname{vec}\left(\mathbf{V}\mathbf{X}_r\mathbf{W}^T\right)^T \operatorname{vec}(\mathbf{G}) \\ &\quad + \operatorname{vec}\left(\mathbf{V}\mathbf{X}_r\mathbf{W}^T\right)^T (\mathbf{M}\otimes\mathbf{A}+\mathbf{H}\otimes\mathbf{E}) \operatorname{vec}\left(\mathbf{V}\mathbf{X}_r\mathbf{W}^T\right) \\ &= \operatorname{tr}\left(\mathbf{X}^T\mathbf{G}\right) - 2\operatorname{tr}\left(\mathbf{W}\mathbf{X}_r^T\mathbf{V}^T\mathbf{G}\right) \\ &\quad + \operatorname{tr}\left(\mathbf{W}\mathbf{X}_r^T\mathbf{V}^T(\mathbf{A}\mathbf{V}\mathbf{X}_r\mathbf{W}^T\mathbf{M}+\mathbf{E}\mathbf{V}\mathbf{X}_r\mathbf{W}^T\mathbf{H})\right). \end{aligned}$$

Using that $\operatorname{tr}(\mathbf{K}) = \operatorname{tr}\left(\mathbf{K}^T\right)$ and $\operatorname{tr}(\mathbf{K}\mathbf{L}) = \operatorname{tr}(\mathbf{L}\mathbf{K})$ for matrices $\mathbf{K},\mathbf{L}$ of compatible dimensions together with (2.3) gives

$$\begin{aligned} \frac{\partial f}{\partial \mathbf{V}} &= 2(\mathbf{A}\mathbf{V}\mathbf{X}_r\mathbf{W}^T\mathbf{M}+\mathbf{E}\mathbf{V}\mathbf{X}_r\mathbf{W}^T\mathbf{H}-\mathbf{G})\mathbf{W}\mathbf{X}_r^T, \\ \frac{\partial f}{\partial \mathbf{W}} &= 2(\mathbf{M}\mathbf{W}\mathbf{X}_r^T\mathbf{V}^T\mathbf{A}\mathbf{V}\mathbf{X}_r+\mathbf{H}\mathbf{W}\mathbf{X}_r^T\mathbf{V}^T\mathbf{E}-\mathbf{G}^T)\mathbf{V}\mathbf{X}_r, \\ \frac{\partial f}{\partial \mathbf{X}_r} &= 2\mathbf{V}^T(\mathbf{A}\mathbf{V}\mathbf{X}_r\mathbf{W}^T\mathbf{M}+\mathbf{E}\mathbf{V}\mathbf{X}_r\mathbf{W}^T\mathbf{H}-\mathbf{G})\mathbf{W}. \end{aligned}$$

Since a minimizer has to satisfy the first-order necessary optimality conditions, it also holds that

$$\frac{\partial f}{\partial \mathbf{V}} = \frac{\partial f}{\partial \mathbf{W}} = \frac{\partial f}{\partial \mathbf{X}_r} = \mathbf{0}.$$

Together with $\mathbf{X}_r$ being nonsingular, this shows the assertion. □

Along the lines of [40], one might consider solving (2.2) by a Riemannian optimization method. While this certainly is possible, in what follows we prefer to proceed via a connection of (2.2) and the $\mathcal{H}_2$-inner product of two dynamical systems. This particularly results in a conceptionally simpler algorithm, which is easy to implement.

3. Tangential interpolation of symmetric state space systems. In this section, it will prove beneficial to assume that the right hand side $\mathbf{G}$ is given in factored form $\mathbf{G} = \mathbf{B}\mathbf{C}^T$ with $\mathbf{B} \in \mathbb{R}^{n\times q}$ and $\mathbf{C} \in \mathbb{R}^{m\times q}$. At this point, it is not particularly important that we have $q \ll n, m$. This also means we can always ensure such a decomposition by, e.g., the SVD of $\mathbf{G}$. We now can associate the energy norm of the solution $\mathbf{X}$ with the $\mathcal{H}_2$-inner product of two dynamical systems defined by their transfer functions. For this, recall that if a symmetric state space system is given as

$$\begin{aligned} \mathbf{E}\dot{\mathbf{x}}(t) &= -\mathbf{A}\mathbf{x}(t)+\mathbf{B}\mathbf{u}(t), \\ \mathbf{y}(t) &= \mathbf{B}^T\mathbf{x}(t), \end{aligned} \tag{3.1}$$

with $\mathbf{x}(t) \in \mathbb{R}^{n\times n}, \mathbf{u}(t), \mathbf{y}(t) \in \mathbb{R}^q$, denoting state, control, and output of the system, the *transfer function* is the rational matrix valued function

$$\mathbf{G}_1(s) = \mathbf{B}^T(s\mathbf{E}+\mathbf{A})^{-1}\mathbf{B}.$$

Since $\mathbf{E}, \mathbf{A} \succ 0$, system (3.1) is asymptotically stable and the poles of $\mathbf{G}_1(s)$ are all in the open left half of the complex plane. Hence, for $\mathbf{G}_1(s)$ and

$$\mathbf{G}_2(s) := \mathbf{C}^T(s\mathbf{M}+\mathbf{H})^{-1}\mathbf{C},$$

ETNA
Kent State University
http://etna.math.kent.edu

the $\mathcal{H}_2$-inner product is defined as

$$\begin{aligned}\langle \mathbf{G}_1, \mathbf{G}_2 \rangle_{\mathcal{H}_2} &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \operatorname{trace}\left(\overline{\mathbf{G}_1(\imath\omega)} \mathbf{G}_2(\imath\omega)^T \right) \mathrm{d}\omega \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \operatorname{trace}\left(\mathbf{G}_1(-\imath\omega) \mathbf{G}_2(\imath\omega)^T \right) \mathrm{d}\omega.\end{aligned}$$

The previous expression turns out to be exactly the square of the energy norm of $\mathbf{X}$.

PROPOSITION 3.1. *Let* $\mathbf{X}$ *be the solution of* $\mathbf{AXM} + \mathbf{EXH} = \mathbf{BC}^T$. *Then it holds that*

$$\|\mathbf{X}\|_{\mathcal{L}_S}^2 = \langle \mathbf{G}_1, \mathbf{G}_2 \rangle_{\mathcal{H}_2},$$

where $\mathbf{G}_1(s) = \mathbf{B}^T(s\mathbf{E} + \mathbf{A})^{-1}\mathbf{B}$ *and* $\mathbf{G}_2(s) = \mathbf{C}^T(s\mathbf{M} + \mathbf{H})^{-1}\mathbf{C}$.

Proof. First note that we have

$$\|\mathbf{X}\|_{\mathcal{L}_S}^2 = \operatorname{vec}(\mathbf{X})^T (\mathbf{M} \otimes \mathbf{A} + \mathbf{H} \otimes \mathbf{E}) \operatorname{vec}(\mathbf{X}).$$

Since $\mathbf{X}$ is a solution of the Sylvester equation, this implies that

$$\|\mathbf{X}\|_{\mathcal{L}_S}^2 = \operatorname{vec}(\mathbf{X})^T \operatorname{vec}\left(\mathbf{BC}^T\right).$$

Due to the properties of the trace-operator, we have

$$\|\mathbf{X}\|_{\mathcal{L}_S}^2 = \operatorname{trace}\left(\mathbf{X}^T\mathbf{BC}^T\right) = \operatorname{trace}\left(\mathbf{B}^T\mathbf{XC}\right).$$

On the other hand, it is well-known (see, e.g., [1]) that the solution of a Sylvester equation can be obtained by complex integration as

$$\mathbf{X} = \frac{1}{2\pi} \int_{-\infty}^{\infty} (-\imath\omega\mathbf{E} + \mathbf{A})^{-1}\mathbf{BC}^T(\imath\omega\mathbf{M} + \mathbf{H})^{-1}\mathrm{d}\omega.$$

Pre- and post-multiplication with, respectively, $\mathbf{B}^T$ and $\mathbf{C}$ show the assertion. □

Instead of parameterizing the minimization problem (2.2) via $\mathbf{V}, \mathbf{W}, \mathbf{X}_r$, the goal is to use *reduced* rational transfer functions

$$\mathbf{G}_{1,r}(s) = \mathbf{B}_r^T(s\mathbf{E}_r + \mathbf{A}_r)^{-1}\mathbf{B}_r \quad \text{and} \quad \mathbf{G}_{2,r}(s) = \mathbf{C}_r^T(s\mathbf{M}_r + \mathbf{H}_r)^{-1}\mathbf{C}_r,$$

with symmetric positive definite matrices $\mathbf{A}_r, \mathbf{E}_r, \mathbf{M}_r$, and $\mathbf{H}_r$ of dimension $n_r \times n_r$ and $\mathbf{B}_r, \mathbf{C}_r \in \mathbb{R}^{n_r \times q}$. Since using every entry of the system matrices would lead to an over-parameterization, we replace $\mathbf{G}_{1,r}$ and $\mathbf{G}_{2,r}$ by their *pole-residue representations*. For this, let $\mathbf{A}_r\mathbf{Q} = \mathbf{E}_r\mathbf{Q}\mathbf{\Lambda}$ be the eigenvalue decomposition of the matrix pencil $(\mathbf{A}_r, \mathbf{E}_r)$. Since $\mathbf{A}_r$ and $\mathbf{E}_r$ are symmetric positive definite, we can choose $\mathbf{Q}^T\mathbf{E}_r\mathbf{Q} = \mathbf{I}$. Hence, we have

$$\mathbf{G}_{1,r}(s) = \mathbf{B}_r^T(s\mathbf{E}_r + \mathbf{A}_r)^{-1}\mathbf{B}_r = \mathbf{B}_r^T\mathbf{Q}(\mathbf{Q}^T(s\mathbf{E}_r + \mathbf{A}_r)\mathbf{Q})^{-1}\mathbf{Q}^T\mathbf{B}_r = \sum_{i=1}^{n_r} \frac{\mathbf{b}_i\mathbf{b}_i^T}{s + \lambda_i},$$

with $\mathbf{\Lambda} = \operatorname{diag}(\lambda_1, \ldots, \lambda_{n_r})$ and $\mathbf{B}_r^T\mathbf{Q} = [\mathbf{b}_1, \ldots, \mathbf{b}_{n_r}]$. The name of the representation is due to the fact that $\mathbf{b}_i\mathbf{b}_i^T = \operatorname{res}[\mathbf{G}_{1,r}(s), \lambda_i]$. Analogously, let $\mathbf{G}_{2,r}(s)$ be given as

$$\mathbf{G}_{2,r}(s) = \sum_{j=1}^{n_r} \frac{\mathbf{c}_j\mathbf{c}_j^T}{s + \sigma_j},$$

ETNA
Kent State University
http://etna.math.kent.edu

where the σ_j are the eigenvalues of the pencil $(\mathbf{H}_r, \mathbf{M}_r)$ and $\mathbf{c}_j\mathbf{c}_j^T = \operatorname{res}[\mathbf{G}_{2,r}(s), \sigma_j]$. Next, define an objective function via

$$\mathcal{J} := \langle \mathbf{G}_1 - \mathbf{G}_{1,r}, \mathbf{G}_2 - \mathbf{G}_{2,r} \rangle_{\mathcal{H}_2}.$$

For reduced transfer functions obtained within a projection framework, in [9] we have claimed that

$$\mathcal{J} \leq \langle \mathbf{G}_1, \mathbf{G}_2 \rangle_{\mathcal{H}_2} - \langle \mathbf{G}_{1,r}, \mathbf{G}_{2,r} \rangle_{\mathcal{H}_2} = \|\mathbf{X} - \tilde{\mathbf{X}}\|_{\mathcal{L}_S}^2,$$

where $\tilde{\mathbf{X}}$ can be constructed by prolongation of the solution $\mathbf{X}_r$ of a reduced Sylvester equation. For the sake of completeness, we give a proof based on the following two results from [2] and [20] (stated here for multi-input multi-output systems).

LEMMA 3.2 [2]. *Suppose that* $\mathbf{G}(s)$ *and* $\mathbf{H}(s) = \sum_{i=1}^{m} \frac{1}{s+\mu_i}\mathbf{c}_i\mathbf{b}_i^T$ *are stable and have simple poles. Then*

$$\langle \mathbf{G}, \mathbf{H} \rangle_{\mathcal{H}_2} = \sum_{i=1}^{m} \mathbf{c}_i^T \overline{\mathbf{G}}(\mu_i)\mathbf{b}_i.$$

LEMMA 3.3 [20]. *Let* $\mathbf{H}(s) = \mathbf{B}^T(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{B}$ *be a symmetric state space system, and let* $\mathbf{H}_r(s) = \mathbf{B}_r^T(s\mathbf{I}_r - \mathbf{A}_r)^{-1}\mathbf{B}_r$ *be any reduced model of* $\mathbf{H}(s)$ *constructed by a compression of* $\mathbf{H}(s)$*, i.e.,* $\mathbf{A}_r = \mathbf{V}^T\mathbf{A}\mathbf{V}, \mathbf{B}_r = \mathbf{V}^T\mathbf{B}$*. Then, for any* $s \geq 0$*,*

$$\mathbf{H}(s) - \mathbf{H}_r(s) \succeq 0.$$

LEMMA 3.4. *Let* $\mathbf{G}_1(s) = \mathbf{B}^T(s\mathbf{E} + \mathbf{A})^{-1}\mathbf{B}$ *and* $\mathbf{G}_2(s) = \mathbf{C}^T(s\mathbf{M} + \mathbf{H})^{-1}\mathbf{C}$ *be given transfer functions. Suppose that* $\mathbf{G}_{1,r}(s) = \mathbf{B}_r^T(s\mathbf{E}_r + \mathbf{A}_r)^{-1}\mathbf{B}_r = \sum_{i=1}^{n_r} \frac{\mathbf{b}_i\mathbf{b}_i^T}{s+\lambda_i}$ *and* $\mathbf{G}_{2,r}(s) = \mathbf{C}_r^T(s\mathbf{M}_r + \mathbf{H}_r)^{-1}\mathbf{C}_r = \sum_{j=1}^{n_r} \frac{\mathbf{c}_j\mathbf{c}_j^T}{s+\sigma_j}$ *have been constructed by orthogonal projections*

$$\begin{aligned} \mathbf{A}_r &= \mathbf{V}^T\mathbf{A}\mathbf{V}, & \mathbf{E}_r &= \mathbf{V}^T\mathbf{E}\mathbf{V}, & \mathbf{B}_r &= \mathbf{V}^T\mathbf{B}, \\ \mathbf{H}_r &= \mathbf{W}^T\mathbf{H}\mathbf{W}, & \mathbf{M}_r &= \mathbf{W}^T\mathbf{M}\mathbf{W}, & \mathbf{C}_r &= \mathbf{W}^T\mathbf{C}. \end{aligned}$$

Then

$$\langle \mathbf{G}_1 - \mathbf{G}_{1,r}, \mathbf{G}_2 - \mathbf{G}_{2,r} \rangle_{\mathcal{H}_2} \leq \langle \mathbf{G}_1, \mathbf{G}_2 \rangle_{\mathcal{H}_2} - \langle \mathbf{G}_{1,r}, \mathbf{G}_{2,r} \rangle_{\mathcal{H}_2}.$$

Proof. For the $\mathcal{H}_2$-inner product, we find

$$\begin{aligned} \langle \mathbf{G}_1 - \mathbf{G}_{1,r}, \mathbf{G}_2 - \mathbf{G}_{2,r} \rangle_{\mathcal{H}_2} = \langle \mathbf{G}_1, \mathbf{G}_2 \rangle_{\mathcal{H}_2} &- \langle \mathbf{G}_{1,r}, \mathbf{G}_2 - \mathbf{G}_{2,r} \rangle_{\mathcal{H}_2} \\ &- \langle \mathbf{G}_{2,r}, \mathbf{G}_1 - \mathbf{G}_{1,r} \rangle_{\mathcal{H}_2} - \langle \mathbf{G}_{1,r}, \mathbf{G}_{2,r} \rangle_{\mathcal{H}_2}. \end{aligned}$$

Applying Lemma 3.2 to the second term gives

$$-\langle \mathbf{G}_{1,r}, \mathbf{G}_2 - \mathbf{G}_{2,r} \rangle_{\mathcal{H}_2} = -\sum_{i=1}^{n_r} \mathbf{b}_i^T \left(\mathbf{G}_2(\lambda_i) - \mathbf{G}_{2,r}(\lambda_i) \right) \mathbf{b}_i.$$

Since $\mathbf{G}_{1,r}$ is constructed by orthogonal projection, it must have stable poles and thus $\lambda_i \geq 0$. Moreover, Lemma 3.3 yields $\mathbf{G}_2(s) - \mathbf{G}_{2,r}(s) \succeq 0$, which shows that

$$-\langle \mathbf{G}_{1,r}, \mathbf{G}_2 - \mathbf{G}_{2,r} \rangle_{\mathcal{H}_2} \leq 0.$$

The same argument yields $\langle \mathbf{G}_{2,r}, \mathbf{G}_1 - \mathbf{G}_{1,r} \rangle_{\mathcal{H}_2} \geq 0$ and proves the statement. ∎

In particular, the proof indicates that equality holds for $\left(\mathbf{G}_2(\lambda_i) - \mathbf{G}_{2,r}(\lambda_i) \right) \mathbf{b}_i = \mathbf{0}$ and $\left(\mathbf{G}_1(\sigma_j) - \mathbf{G}_{1,r}(\sigma_j) \right) \mathbf{c}_j = \mathbf{0}$. Again this generalizes our SISO formulation in [9]. Moreover, the latter condition is directly related to the gradient of $\mathcal{J}$ with respect to the parameters $\mathbf{b}_i, \lambda_i, \mathbf{c}_i$, and σ_i.

ETNA
Kent State University
http://etna.math.kent.edu

THEOREM 3.5. *Let $\mathbf{G}_1(s), \mathbf{G}_2(s), \mathbf{G}_{1,r}(s)$, and $\mathbf{G}_{2,r}(s)$ be symmetric state space systems with simple poles. Suppose that $\lambda_1, \ldots, \lambda_{n_r}$ and $\sigma_1, \ldots, \sigma_{n_r}$ are the poles of the reduced transfer functions with $\operatorname{res}[\mathbf{G}_{1,r}(s), \lambda_i] = \mathbf{b}_i\mathbf{b}_i^T$ and $\operatorname{res}[\mathbf{G}_{2,r}(s), \sigma_j] = \mathbf{c}_j\mathbf{c}_j^T$, for $i, j = 1, \ldots, n_r$. The gradient of $\mathcal{J}$ with respect to the parameters listed as*

$$\{\mathbf{b}, \boldsymbol{\lambda}, \mathbf{c}, \boldsymbol{\sigma}\} = [\mathbf{b}_1^T, \lambda_1, \mathbf{c}_1^T, \sigma_1, \ldots, \mathbf{b}_{n_r}^T, \lambda_{n_r}, \mathbf{c}_{n_r}^T, \sigma_{n_r}]^T$$

is given by $\nabla_{\{\mathbf{b},\boldsymbol{\lambda},\mathbf{c},\boldsymbol{\sigma}\}}\mathcal{J}$, a vector of length $2n_r(q+1)$ partitioned into n_r vectors of length $2(q+1)$ of the form

$$\left(\nabla_{\{\mathbf{b},\boldsymbol{\lambda},\mathbf{c},\boldsymbol{\sigma}\}}\mathcal{J}\right)_k = \begin{bmatrix} 2\left(\mathbf{G}_{2,r}(\lambda_k) - \mathbf{G}_2(\lambda_k)\right)\mathbf{b}_k \\ \mathbf{b}_k^T(\mathbf{G}'_{2,r}(\lambda_k) - \mathbf{G}'_2(\lambda_k))\mathbf{b}_k \\ 2\left(\mathbf{G}_{1,r}(\sigma_k) - \mathbf{G}_1(\sigma_k)\right)\mathbf{c}_k \\ \mathbf{c}_k^T(\mathbf{G}'_{1,r}(\sigma_k) - \mathbf{G}'_1(\sigma_k))\mathbf{c}_k \end{bmatrix},$$

for $k = 1, \ldots, n_r$.

Proof. Observe that for the ℓ-th entry of $\mathbf{b}_k$, we have

$$\begin{aligned} \frac{\partial\mathcal{J}}{\partial(\mathbf{b}_k)_\ell} &= \frac{\partial}{\partial(\mathbf{b}_k)_\ell}\langle\mathbf{G}_1 - \mathbf{G}_{1,r}, \mathbf{G}_2 - \mathbf{G}_{2,r}\rangle_{\mathcal{H}_2} = -\left\langle\frac{\partial\mathbf{G}_{1,r}}{\partial(\mathbf{b}_k)_\ell}, \mathbf{G}_2 - \mathbf{G}_{2,r}\right\rangle_{\mathcal{H}_2} \\ &= -\left\langle\frac{\mathbf{e}_\ell\mathbf{b}_k^T}{s+\lambda_k}, \mathbf{G}_2 - \mathbf{G}_{2,r}\right\rangle_{\mathcal{H}_2} - \left\langle\frac{\mathbf{b}_k\mathbf{e}_\ell^T}{s+\lambda_k}, \mathbf{G}_2 - \mathbf{G}_{2,r}\right\rangle_{\mathcal{H}_2} \\ &= -\mathbf{e}_\ell^T\left(\mathbf{G}_2(\lambda_k) - \mathbf{G}_{2,r}(\lambda_k)\right)\mathbf{b}_k - \mathbf{b}_k^T\left(\mathbf{G}_2(\lambda_k) - \mathbf{G}_{2,r}(\lambda_k)\right)\mathbf{e}_\ell \\ &= -2\mathbf{e}_\ell^T\left(\mathbf{G}_2(\lambda_k) - \mathbf{G}_{2,r}(\lambda_k)\right)\mathbf{b}_k, \end{aligned}$$

where $\mathbf{e}_\ell$ is the ℓ-th unit vector. The previous steps follow from Lemma 3.2 and the fact that $\mathbf{G}_2$ and $\mathbf{G}_{2,r}$ are symmetric state space systems. Similarly, for the derivative with respect to λ_k, we find

$$\begin{aligned} \frac{\partial\mathcal{J}}{\partial\lambda_k} &= \frac{\partial}{\partial\lambda_k}\langle\mathbf{G}_1 - \mathbf{G}_{1,r}, \mathbf{G}_2 - \mathbf{G}_{2,r}\rangle_{\mathcal{H}_2} = -\left\langle\frac{\partial}{\partial\lambda_k}\mathbf{G}_{1,r}, \mathbf{G}_2 - \mathbf{G}_{2,r}\right\rangle_{\mathcal{H}_2} \\ &= \left\langle\frac{\mathbf{b}_k\mathbf{b}_k^T}{(s+\lambda_k)^2}, \mathbf{G}_2 - \mathbf{G}_{2,r}\right\rangle_{\mathcal{H}_2}. \end{aligned}$$

For the latter expression, we can use the MIMO analogue of [27, Lemma 2.4] and obtain

$$\frac{\partial\mathcal{J}}{\partial\lambda_k} = \mathbf{b}_k^T(\mathbf{G}'_{2,r}(\lambda_k) - \mathbf{G}'_2(\lambda_k))\mathbf{b}_k.$$

The proofs for $\mathbf{c}_k$ and σ_k use exactly the same arguments and are thus omitted here. ∎

REMARK 3.6. Note the change of sign for the derivatives with respect to λ_k and σ_k compared to the special case of $\mathcal{H}_2$-optimal model reduction discussed in [7]. This simply follows from a different notation in this manuscript. Using $\lambda_i, \sigma_j < 0$ together with transfer function representations $\sum_{i=1}^{n_r}\mathbf{G}_{1,r}(s) = \frac{\mathbf{b}_i\mathbf{b}_i^T}{s-\lambda_i}$ and $\mathbf{G}_{r,2}(s) = \sum_{j=1}^{n_r}\frac{\mathbf{c}_j\mathbf{c}_j^T}{s-\sigma_j}$ would lead to similar expressions as in [7].

In [9], we stated the inequality from Lemma 3.4 and showed that equality holds if the gradient of $\mathcal{J}$ is zero. In fact, we can even show that the corresponding reduced transfer

ETNA
Kent State University
http://etna.math.kent.edu

functions can be used to compute a triple $(\mathbf{V}, \mathbf{W}, \mathbf{X}_r)$ satisfying the first-order necessary optimality conditions from Theorem 2.1.

THEOREM 3.7. *Consider the Sylvester equation* (2.1) *with factored right-hand side* $\mathbf{G} = \mathbf{BC}$,

$$\mathbf{AXM} + \mathbf{EXH} = \mathbf{BC},$$

and denote, respectively, $\mathbf{G}_1(s) = \mathbf{B}^T(s\mathbf{E} + \mathbf{A})^{-1}\mathbf{B}$ *and* $\mathbf{G}_2(s) = \mathbf{C}^T(s\mathbf{M} + \mathbf{H})^{-1}\mathbf{C}$. *Suppose* $\mathbf{G}_{1,r}(s) = \sum_{i=1}^{n_r} \frac{\mathbf{b}_i\mathbf{b}_i^T}{s+\lambda_i}$ *and* $\mathbf{G}_{2,r}(s) = \sum_{j=1}^{n_r} \frac{\mathbf{c}_j\mathbf{c}_j^T}{s+\sigma_j}$ *satisfy*

$$\mathbf{G}_{1,r}(\sigma_k)\mathbf{c}_k = \mathbf{G}_1(\sigma_k)\mathbf{c}_k, \tag{3.2a}$$

$$\mathbf{c}_k^T\mathbf{G}'_{1,r}(\sigma_k)\mathbf{c}_k = \mathbf{c}_k^T\mathbf{G}'_1(\sigma_k)\mathbf{c}_k, \tag{3.2b}$$

$$\mathbf{G}_{2,r}(\lambda_k)\mathbf{b}_k = \mathbf{G}_2(\lambda_k)\mathbf{b}_k, \tag{3.2c}$$

$$\mathbf{b}_k^T\mathbf{G}'_{2,r}(\lambda_k)\mathbf{b}_k = \mathbf{b}_k^T\mathbf{G}'_2(\lambda_k)\mathbf{b}_k, \tag{3.2d}$$

for $k = 1, \ldots, n_r$. *Define* $\mathbb{X} \in \mathbb{R}^{n_r \times n_r}$, $\mathbb{Y} \in \mathbb{R}^{n \times n_r}$, *and* $\mathbb{Z} \in \mathbb{R}^{m \times n_r}$ *via*

$$\mathbb{X}_{ij} = \frac{\mathbf{b}_i^T\mathbf{c}_j}{\lambda_i + \sigma_j}, \quad \mathbb{Y}_i = (\sigma_i\mathbf{E} + \mathbf{A})^{-1}\mathbf{Bc}_i, \quad \mathbb{Z}_j = (\lambda_j\mathbf{M} + \mathbf{H})^{-1}\mathbf{Cb}_j.$$

Then the triple $(\mathbb{Y}, \mathbb{Z}, \mathbb{X}^{-1})$ *satisfies* (2.4).

Proof. First note that (3.2) defines $n_r(q+1)$ constraints on, respectively, $\mathbf{G}_{1,r}(s)$ and $\mathbf{G}_{2,r}(s)$. Due to the pole-residue representation, exactly the same number of parameters defines the rational matrix valued transfer functions $\mathbf{G}_{1,r}(s)$ and $\mathbf{G}_{2,r}(s)$. Hence, $\mathbf{G}_{1,r}(s)$ and $\mathbf{G}_{2,r}$ are uniquely determined by (3.2). Echoing the argumentation in [17, Lemma 3.11] and [34], without loss of generality we can thus assume that the reduced transfer functions are obtained by $\mathbf{E}/\mathbf{M}$-orthogonal projections via

$$\mathbf{\Lambda} = \mathbf{V}^T\mathbf{AV}, \qquad \tilde{\mathbf{B}} := [\mathbf{b}_1, \ldots, \mathbf{b}_q]^T = \mathbf{V}^T\mathbf{B},$$
$$\mathbf{\Sigma} = \mathbf{W}^T\mathbf{HW}, \qquad \tilde{\mathbf{C}} := [\mathbf{c}_1, \ldots, \mathbf{c}_q]^T = \mathbf{W}^T\mathbf{C},$$

where $\mathbf{V}$ and $\mathbf{W}$ are such that

$$\operatorname{span}\{\mathbf{V}\} \supset \operatorname*{span}_{i=1,\ldots,n_r} \left\{(\sigma_i\mathbf{E} + \mathbf{A})^{-1}\mathbf{Bc}_i\right\},$$
$$\operatorname{span}\{\mathbf{W}\} \supset \operatorname*{span}_{j=1,\ldots,n_r} \left\{(\lambda_j\mathbf{M} + \mathbf{H})^{-1}\mathbf{Cb}_j\right\}.$$

Due to the definition of $\mathbb{X}_{ij}$ we further obtain

$$\mathbb{X}_i = (\sigma_i\mathbf{I} + \mathbf{\Lambda})^{-1}\tilde{\mathbf{B}}\mathbf{c}_i, \quad \mathbb{X}_j^T = (\lambda_j\mathbf{I} + \mathbf{\Sigma})^{-1}\tilde{\mathbf{C}}\mathbf{b}_j,$$

where $\mathbf{\Lambda} = \operatorname{diag}(\lambda_1, \ldots, \lambda_{n_r})$ and $\mathbf{\Sigma} = \operatorname{diag}(\sigma_1, \ldots, \sigma_{n_r})$. Using well-known results from projection-based rational interpolation (see [26]), we conclude

$$\mathbf{V}\mathbb{X}_i = \mathbb{Y}_i, \quad \mathbf{W}\mathbb{X}_j^T = \mathbb{Z}_j,$$

and therefore $\mathbf{V}\mathbb{X} = \mathbb{Y}$ and $\mathbf{W}\mathbb{X}^T = \mathbb{Z}$. Keeping this in mind, for (2.4), we obtain

$$\begin{aligned}
&\left(\mathbf{A}\mathbb{Y}\mathbb{X}^{-1}\mathbb{Z}^T\mathbf{M} + \mathbf{E}\mathbb{Y}\mathbb{X}^{-1}\mathbb{Z}^T\mathbf{H} - \mathbf{BC}^T\right)\mathbb{Z} \\
&\quad = \left(\mathbf{A}\mathbb{Y}\mathbf{W}^T\mathbf{M} + \mathbf{E}\mathbb{Y}\mathbf{W}^T\mathbf{H} - \mathbf{BC}^T\right)\mathbf{W}\mathbb{X}^T \\
&\quad = (\mathbf{A}\mathbb{Y} + \mathbf{E}\mathbb{Y}\mathbf{\Sigma} - \mathbf{B}\tilde{\mathbf{C}}^T)\mathbb{X}^T = \mathbf{0}.
\end{aligned}$$

ETNA
Kent State University
http://etna.math.kent.edu

Here, the last step follows from the definition of $\mathbb{Y}$. Similarly, it holds that

$$\begin{aligned}\mathbb{Y}^T\left(\mathbf{A}\mathbb{Y}\mathbb{X}^{-1}\mathbb{Z}^T\mathbf{M}+\mathbf{E}\mathbb{Y}\mathbb{X}^{-1}\mathbb{Z}^T\mathbf{H}-\mathbf{B}\mathbf{C}^T\right)&=\mathbb{X}^T\mathbf{V}^T\left(\mathbf{A}\mathbf{V}\mathbb{Z}^T\mathbf{M}+\mathbf{E}\mathbf{V}\mathbb{Z}^T\mathbf{H}-\mathbf{B}\mathbf{C}^T\right)\\&=\mathbb{X}^T(\mathbf{\Lambda}\mathbb{Z}^T\mathbf{M}+\mathbb{Z}^T\mathbf{H}-\tilde{\mathbf{B}}\mathbf{C}^T)=\mathbf{0}.\end{aligned}$$

Again, the last equality is due to the definition of $\mathbb{Z}$. Finally, we have

$$\begin{aligned}&\mathbb{Y}^T\left(\mathbf{A}\mathbb{Y}\mathbb{X}^{-1}\mathbb{Z}^T\mathbf{M}+\mathbf{E}\mathbb{Y}\mathbb{X}^{-1}\mathbb{Z}^T\mathbf{H}-\mathbf{B}\mathbf{C}^T\right)\mathbb{Z}\\&\quad=\mathbb{X}^T\mathbf{V}^T\left(\mathbf{A}\mathbf{V}\mathbb{X}\mathbf{W}^T\mathbf{M}+\mathbf{E}\mathbf{V}\mathbb{X}\mathbf{W}^T\mathbf{H}-\mathbf{B}\mathbf{C}^T\right)\mathbf{W}\mathbb{X}^T\\&\quad=\mathbb{X}^T\left(\mathbf{\Lambda}\mathbb{X}+\mathbb{X}\mathbf{\Sigma}-\tilde{\mathbf{B}}\tilde{\mathbf{C}}^T\right)\mathbb{X}^T=\mathbf{0}.\end{aligned}$$

Once more, the last identity is true due to the definition of $\mathbb{X}$. ∎

REMARK 3.8. From the proof of Theorem 3.7, we find that the same approximation is obtained when $(\mathbb{Y},\mathbb{Z},\mathbb{X}^{-1})$ is replaced by $(\mathbf{V},\mathbf{W},\mathbb{X})$ where $\mathbf{V}$ and $\mathbf{W}$ are the projection matrices constructing $\mathbf{G}_{1,r}(s)$ and $\mathbf{G}_{2,r}(s)$. Furthermore note that $\mathbb{X}$ solves the projected reduced Sylvester equation. This in particular implies that the approximation $\mathbf{V}\mathbb{X}\mathbf{W}^T$ fulfills the common Galerkin condition on the residual; see [37].

The natural question that arises is whether triples $(\mathbf{V},\mathbf{W},\mathbf{X}_r)$ fulfilling (2.4) also yield reduced transfer functions $\mathbf{G}_{1,r}(s)$ and $\mathbf{G}_{2,r}(s)$ with vanishing gradient $\nabla_{\{\mathbf{b},\boldsymbol{\lambda},\mathbf{c},\boldsymbol{\sigma}\}}\mathcal{J}$. The answer is given by the following result.

THEOREM 3.9. *Let a triple $(\mathbf{V},\mathbf{W},\mathbf{X}_r)$ be given such that (2.4) holds. Suppose reduced transfer functions $\mathbf{G}_{1,r}(s)$ and $\mathbf{G}_{2,r}(s)$ are defined via*

$$\begin{aligned}\mathbf{A}_r&=\mathbf{V}^T\mathbf{A}\mathbf{V}, & \mathbf{E}_r&=\mathbf{V}^T\mathbf{E}\mathbf{V}, & \mathbf{B}_r&=\mathbf{V}^T\mathbf{B},\\ \mathbf{H}_r&=\mathbf{W}^T\mathbf{H}\mathbf{W}, & \mathbf{M}_r&=\mathbf{W}^T\mathbf{M}\mathbf{W}, & \mathbf{C}_r&=\mathbf{W}^T\mathbf{C}.\end{aligned}$$

Then it holds that $\nabla_{\{\mathbf{b},\boldsymbol{\lambda},\mathbf{c},\boldsymbol{\sigma}\}}\mathcal{J}=\mathbf{0}$.

Proof. The third condition in (2.4) implies

$$\mathbf{A}_r\mathbf{X}_r\mathbf{M}_r+\mathbf{E}_r\mathbf{X}_r\mathbf{H}_r-\mathbf{B}_r\mathbf{C}_r^T=\mathbf{0}.$$

Assuming that $\mathbf{H}_r\mathbf{R}=\mathbf{M}_r\mathbf{R}\mathbf{\Sigma}$ is the eigenvalue decomposition of $(\mathbf{H}_r,\mathbf{M}_r)$, post-multiplication of the above equation with $\mathbf{r}_j:=\mathbf{R}\mathbf{e}_j$ gives

$$\mathbf{A}_r\underbrace{\mathbf{X}_r\mathbf{M}_r\mathbf{r}_j}_{\mathbf{x}_j}+\sigma_j\mathbf{E}_r\underbrace{\mathbf{X}_r\mathbf{M}_r\mathbf{r}_j}_{\mathbf{x}_j}=\mathbf{B}_r\underbrace{\mathbf{C}_r^T\mathbf{r}_j}_{\mathbf{c}_j}.$$

Hence, we have $\mathbf{x}_j=(\sigma_j\mathbf{E}_r+\mathbf{A}_r)^{-1}\mathbf{B}_r\mathbf{c}_j$. Also, post-multiplication of the third equation in (2.4) with $\mathbf{r}_j$ yields

$$\mathbf{A}\mathbf{V}\mathbf{X}_r\mathbf{M}_r\mathbf{r}_j+\sigma_j\mathbf{E}\mathbf{V}\mathbf{X}_r\mathbf{M}_r\mathbf{r}_j=\mathbf{B}\mathbf{C}_r^T\mathbf{r}_j.$$

In particular, we conclude $\mathbf{V}\mathbf{x}_j=(\sigma\mathbf{E}+\mathbf{A})^{-1}\mathbf{B}\mathbf{c}_j$. This, however, yields

$$\begin{aligned}\mathbf{G}_{1,r}(\sigma_j)\mathbf{c}_j&=\mathbf{B}^T\mathbf{V}(\sigma_j\mathbf{E}_r+\mathbf{A}_r)^{-1}\mathbf{B}\mathbf{c}_j=\mathbf{B}^T(\sigma_j\mathbf{E}+\mathbf{A})^{-1}\mathbf{B}\mathbf{c}_j=\mathbf{G}_1(\sigma_j)\mathbf{c}_j,\\ \mathbf{c}_j^T\mathbf{G}_{1,r}'(\sigma_j)\mathbf{c}_j&=-\mathbf{c}_j^T\mathbf{B}_r^T(\sigma_j\mathbf{E}_r+\mathbf{A}_r)^{-1}\mathbf{V}^T\mathbf{E}\mathbf{V}(\sigma_j\mathbf{E}_r+\mathbf{A}_r)^{-1}\mathbf{B}_r\mathbf{c}_j\\&=-\mathbf{c}_j^T\mathbf{B}^T(\sigma_j\mathbf{E}+\mathbf{A})^{-1}\mathbf{E}(\sigma_j\mathbf{E}+\mathbf{A})^{-1}\mathbf{B}\mathbf{c}_j=\mathbf{c}_j^T\mathbf{G}_1'(\sigma_j)\mathbf{c}_j.\end{aligned}$$

The proof for $\mathbf{G}_{2,r}$ follows analogously. ∎

ETNA
Kent State University
http://etna.math.kent.edu

In summary, we can state that the first-order necessary optimality conditions for the objective functions $f(\mathbf{V},\mathbf{W},\mathbf{X}_r)$ and $\mathcal{J}(\mathbf{b},\boldsymbol{\lambda},\mathbf{c},\boldsymbol{\sigma})$ are equivalent to each other. For the remainder of this paper, we focus on the objective function $\mathcal{J}$. Along the lines of [7], we present the Hessian of $\mathcal{J}$ with respect to the parameters $\{\mathbf{b},\boldsymbol{\lambda},\mathbf{c},\boldsymbol{\sigma}\}$.

LEMMA 3.10. *The Hessian of $\mathcal{J}$ with respect to $\{\mathbf{b},\boldsymbol{\lambda},\mathbf{c},\boldsymbol{\sigma}\}$ is given by $\nabla^2_{\{\mathbf{b},\boldsymbol{\lambda},\mathbf{c},\boldsymbol{\sigma}\}}\mathcal{J}$, a $(2n_r(q+1))\times(2n_r(q+1))$ matrix partitioned into n_r^2 matrices of size $2(q+1)\times 2(q+1)$ defined by*

$$\left(\nabla^2_{\{\mathbf{b},\boldsymbol{\lambda},\mathbf{c},\boldsymbol{\sigma}\}}\mathcal{J}\right)_{k\ell}$$

$$= \begin{bmatrix} \mathbf{0} & \mathbf{0} & 2\left(\frac{\mathbf{c}_\ell\mathbf{b}_k^T+\mathbf{c}_\ell^T\mathbf{b}_k\mathbf{I}_q}{\sigma_\ell+\lambda_k}\right) & -2\frac{\mathbf{c}_\ell\mathbf{c}_\ell^T\mathbf{b}_k}{(\sigma_\ell+\lambda_k)^2} \\ \mathbf{0} & 0 & -2\frac{\mathbf{c}_\ell^T\mathbf{b}_k\mathbf{b}_k^T}{(\lambda_k+\sigma_\ell)^2} & 2\frac{\mathbf{b}_k^T\mathbf{c}_\ell\mathbf{c}_\ell^T\mathbf{b}_k}{(\sigma_\ell+\lambda_k)^3} \\ 2\left(\frac{\mathbf{b}_\ell\mathbf{c}_k^T+\mathbf{b}_\ell^T\mathbf{c}_k\mathbf{I}_q}{\sigma_k+\lambda_\ell}\right) & -2\frac{\mathbf{b}_\ell\mathbf{b}_\ell^T\mathbf{c}_k}{(\sigma_k+\lambda_\ell)^2} & \mathbf{0} & \mathbf{0} \\ -2\frac{\mathbf{b}_\ell^T\mathbf{c}_k\mathbf{c}_k^T}{(\lambda_\ell+\sigma_k)^2} & 2\frac{\mathbf{c}_k^T\mathbf{b}_\ell\mathbf{b}_\ell^T\mathbf{c}_k}{(\lambda_\ell+\sigma_k)^3} & \mathbf{0} & 0 \end{bmatrix}$$

$$+\delta_{k\ell}\begin{bmatrix} 2(\mathbf{G}_{2,r}(\lambda_k)-\mathbf{G}_2(\lambda_k)) & 2(\mathbf{G}'_{2,r}(\lambda_k)-\mathbf{G}'_2(\lambda_k))\mathbf{b}_k & \mathbf{0} & \mathbf{0} \\ 2\mathbf{b}_k^T(\mathbf{G}'_{2,r}(\lambda_k)-\mathbf{G}'_2(\lambda_k)) & \mathbf{b}_k^T\left(\mathbf{G}''_{2,r}(\lambda_k)-\mathbf{G}''_2(\lambda_k)\right)\mathbf{b}_k & \mathbf{0} & 0 \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & 0 & \mathbf{0} & 0 \end{bmatrix}$$

$$+\delta_{k\ell}\begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & 0 & \mathbf{0} & 0 \\ \mathbf{0} & \mathbf{0} & 2(\mathbf{G}_{1,r}(\sigma_k)-\mathbf{G}_1(\sigma_k)) & 2(\mathbf{G}'_{1,r}(\sigma_k)-\mathbf{G}'_1(\sigma_k))\mathbf{c}_k \\ \mathbf{0} & 0 & 2\mathbf{c}_k^T(\mathbf{G}'_{1,r}(\sigma_k)-\mathbf{G}'_1(\sigma_k)) & \mathbf{c}_k^T\left(\mathbf{G}''_{1,r}(\sigma_k)-\mathbf{G}''_1(\sigma_k)\right)\mathbf{c}_k \end{bmatrix}.$$

The proof follows by direct computation of the partial derivatives. Since a similar derivation can be found in [7] for the $\mathcal{H}_2$-optimal case, we omit the details.

Unfortunately, the objective function $\mathcal{J}$ is unbounded so that its minimization is not well defined. This can be seen by considering $n_r = 1$. In this case,

$$\mathbf{G}_{1,r}(s) = \frac{\mathbf{b}\mathbf{b}^T}{s+\lambda} \quad \text{and} \quad \mathbf{G}_{2,r}(s) = \frac{\mathbf{c}\mathbf{c}^T}{s+\mu}$$

are the reduced transfer functions. By Lemma 3.2, for the objective function we get

$$\mathcal{J} = \langle \mathbf{G}_1,\mathbf{G}_2\rangle_{\mathcal{H}_2} - \mathbf{b}^T\mathbf{G}_2(\lambda)\mathbf{b} - \mathbf{c}^T\mathbf{G}_1(\mu)\mathbf{c} + \frac{\mathbf{b}^T\mathbf{c}\mathbf{c}^T\mathbf{b}}{\lambda+\mu}.$$

Hence, by scaling $\alpha\mathbf{b}$ and $\frac{1}{\alpha}\mathbf{c}$, we further obtain

$$\mathcal{J} = \langle \mathbf{G}_1,\mathbf{G}_2\rangle_{\mathcal{H}_2} - \alpha^2\mathbf{b}^T\mathbf{G}_2(\lambda)\mathbf{b} - \frac{1}{\alpha^2}\mathbf{c}^T\mathbf{G}_1(\mu)\mathbf{c} + \frac{\mathbf{b}^T\mathbf{c}\mathbf{c}^T\mathbf{b}}{\lambda+\mu},$$

and we can arbitrarily decrease the value of $\mathcal{J}$ by increasing α. In fact, a similar conclusion can be drawn from the Hessian in Theorem 3.10. Multiplication of $\left(\nabla^2_{\{\mathbf{b},\boldsymbol{\lambda},\mathbf{c},\boldsymbol{\sigma}\}}\mathcal{J}\right)_{11}$ with $\mathbf{z} := \begin{bmatrix}\alpha\mathbf{b}_1^T & 0 & \mathbf{c}_1^T & 0\end{bmatrix}^T$ yields

$$\mathbf{z}^T\left(\nabla^2_{\{\mathbf{b},\boldsymbol{\lambda},\mathbf{c},\boldsymbol{\sigma}\}}\mathcal{J}\right)_{11}\mathbf{z} = 2\alpha^2\mathbf{b}_1^T(\mathbf{G}_{2,r}(\lambda_1)-\mathbf{G}_2(\lambda_1))\mathbf{b}_1 + 2\mathbf{c}_1^T(\mathbf{G}_{1,r}(\sigma_1)-\mathbf{G}_2(\sigma_1))\mathbf{c}_1 + 8\alpha\frac{(\mathbf{b}_1^T\mathbf{c}_1)^2}{\sigma_1+\lambda_1}.$$

ETNA
Kent State University
http://etna.math.kent.edu

For a stationary point, we thus find

$$\mathbf{z}^T \left(\nabla^2_{\{\mathbf{b},\boldsymbol{\lambda},\mathbf{c},\boldsymbol{\sigma}\}} \mathcal{J} \right)_{11} \mathbf{z} = 8\alpha \frac{(\mathbf{b}_1^T \mathbf{c}_1)^2}{\sigma_1 + \lambda_1}.$$

In other words, the Hessian is always indefinite and, consequently, all stationary points are saddle points. While this will cause problems for optimization routines, we can still extend the idea of iterative correction as in [27] to the MIMO Sylvester case. Algorithm 1 is a suitable generalization of a SISO version we proposed in [9]. Due to the iterative structure, upon convergence, the reduced transfer functions $\mathbf{G}_{1,r}(s)$ and $\mathbf{G}_{2,r}(s)$ will tangentially interpolate the original transfer function $\mathbf{G}_1(s)$ and $\mathbf{G}_2(s)$ such that the corresponding gradient in Lemma 3.5 vanishes. According to Theorem 3.7, in this way we can compute stationary points of the objective function f, which is obviously bounded.

ALGORITHM 1: MIMO (Sy)^{2}IRKA

Input: Interpolation points: $\{\lambda_1, \ldots, \lambda_{n_r}\}$ and $\{\sigma_1, \ldots, \sigma_{n_r}\}$.
Tangential directions: $\tilde{\mathbf{B}} = [\mathbf{b}_1, \ldots, \mathbf{b}_{n_r}]$ and $\tilde{\mathbf{C}} = [\mathbf{c}_1, \ldots, \mathbf{c}_{n_r}]$.
Output: $\mathbf{G}_{1,r}(s), \mathbf{G}_{2,r}(s)$ satisfying (3.2)

1: **while** relative change in $\{\lambda_i, \sigma_i\} >$ tol **do**
2: Compute $\mathbf{V}$ and $\mathbf{W}$ from

$$\operatorname{span}\{\mathbf{V}\} \supset \operatorname*{span}_{i=1,\ldots,n_r} \left\{ (\sigma_i \mathbf{E} + \mathbf{A})^{-1} \mathbf{B} \mathbf{c}_i \right\},$$
$$\operatorname{span}\{\mathbf{W}\} \supset \operatorname*{span}_{j=1,\ldots,n_r} \left\{ (\lambda_j \mathbf{M} + \mathbf{H})^{-1} \mathbf{C} \mathbf{b}_j \right\}.$$

3: Compute $\mathbf{E}_r = \mathbf{V}^T \mathbf{E} \mathbf{V}$, $\mathbf{A}_r = \mathbf{V}^T \mathbf{A} \mathbf{V}$, $\mathbf{B}_r = \mathbf{V}^T \mathbf{B}$.
4: Compute $\mathbf{M}_r = \mathbf{W}^T \mathbf{M} \mathbf{W}$, $\mathbf{H}_r = \mathbf{W}^T \mathbf{H} \mathbf{W}$, $\mathbf{C}_r = \mathbf{W}^T \mathbf{C}$.
5: Compute $\mathbf{A}_r \mathbf{Q} = \mathbf{E}_r \mathbf{Q} \boldsymbol{\Lambda}$ with $\mathbf{Q}^T \mathbf{E}_r \mathbf{Q} = \mathbf{I}$.
6: Compute $\mathbf{H}_r \mathbf{R} = \mathbf{M}_r \mathbf{R} \boldsymbol{\Sigma}$ with $\mathbf{R}^T \mathbf{M}_r \mathbf{R} = \mathbf{I}$.
7: Update $\lambda_i = \operatorname{diag}(\boldsymbol{\Lambda})$, $\tilde{\mathbf{B}} = \mathbf{B}_r^T \mathbf{Q}$, $\sigma_i = \operatorname{diag}(\boldsymbol{\Sigma})$, $\tilde{\mathbf{C}} = \mathbf{C}_r^T \mathbf{R}$.
8: **end while**
9: Set $\mathbf{G}_{1,r}(s) = \mathbf{B}_r^T (s\mathbf{E}_r + \mathbf{A}_r)^{-1} \mathbf{B}_r$.
10: Set $\mathbf{G}_{2,r}(s) = \mathbf{C}_r^T (s\mathbf{M}_r + \mathbf{H}_r)^{-1} \mathbf{C}_r$.

3.1. Initialization. The efficiency of Algorithm 1 obviously depends on the number of iterations needed until a typical convergence criterion is satisfied. Hence, an important point is the initialization of the algorithm. Several strategies for choosing interpolation points and tangential directions are possible. However, there exists a natural choice for the applications that we consider in the next section. Below, we will see that a blurred and noisy image sometimes is given as the right hand side $\mathbf{G} = \mathbf{B}\mathbf{C}^T$. Though $\mathbf{G}$ deviates from the original unperturbed image, it still is related to it. In other words, $\mathbf{G}$ can be seen as a (rough) approximation to the solution $\mathbf{X}$ of the underlying Sylvester equation. For this reason, if we are interested in constructing an approximation of rank n_r, we propose to use a truncated singular value decomposition of $\mathbf{G} \approx \mathbf{U}_{n_r} \mathbf{D}_{n_r} \mathbf{Z}_{n_r}^T$, with $\mathbf{U}_{n_r} \in \mathbb{R}^{n \times n_r}$, $\mathbf{Z} \in \mathbb{R}^{m \times n_r}$, and $\mathbf{D}_{n_r} \in \mathbb{R}^{n_r \times n_r}$. Since $\mathbf{U}_{n_r}^T \mathbf{U}_{n_r} = \mathbf{I}$ and $\mathbf{Z}_{n_r}^T \mathbf{Z}_{n_r} = \mathbf{I}$, we can construct an initial reduced model via

$$\mathbf{A}_r = \mathbf{U}_{n_r}^T \mathbf{A} \mathbf{U}_{n_r}, \qquad \mathbf{E}_r = \mathbf{U}_{n_r}^T \mathbf{E} \mathbf{U}_{n_r}, \qquad \mathbf{B}_r = \mathbf{U}_{n_r}^T \mathbf{B},$$
$$\mathbf{H}_r = \mathbf{Z}_{n_r}^T \mathbf{H} \mathbf{Z}_{n_r}, \qquad \mathbf{M}_r = \mathbf{Z}_{n_r}^T \mathbf{M} \mathbf{Z}_{n_r}, \qquad \mathbf{C}_r = \mathbf{Z}_{n_r}^T \mathbf{C}.$$

ETNA
Kent State University
http://etna.math.kent.edu

Initial interpolation points and tangential directions then can be obtained by computing the pole-residue representations for

$$\mathbf{G}_{1,r}(s) = \mathbf{B}_r^T(s\mathbf{E}_r + \mathbf{A}_r)^{-1}\mathbf{B}_r \quad \text{and} \quad \mathbf{G}_{2,r}(s) = \mathbf{C}_r^T(s\mathbf{M}_r + \mathbf{H}_r)^{-1}\mathbf{C}_r.$$

In all our numerical examples, we initialize Algorithm 1 by this procedure. Moreover, as we mentioned earlier, the right hand side $\mathbf{G}$ is not necessarily low-rank, and we thus have to face transfer functions with a large number of inputs and outputs. In the case of $\mathcal{H}_2$-optimal model reduction, this can slow down the convergence of iterative algorithms such as IRKA significantly; see [7]. For this reason, in our examples we replace $\mathbf{G}$ by its truncated singular value decomposition, which is of rank n_r. While this means we are actually approximating the solution of a perturbed Sylvester equation, we will see that this does not seem to influence the quality of restored images using this procedure as explained in the next section.

4. Numerical results. We study the performance of Algorithm 1 for two examples from image restoration. At this point, we emphasize that what follows should only be understood as a numerical validation of Algorithm 1. Moreover, due to the dedication of this special issue, we believe that the following examples are particularly appropriate. We are aware of the fact that using matrix equations within image restoration problems is *not state-of-the-art*. Nowadays, methods based on total (generalized) variation and L_1-norm minimization usually produce much more accurate results.

All simulations were generated on an Intel®Core™i5-3317U CPU, 3 GB RAM, Ubuntu Linux 12.10, MATLAB® Version 7.14.0.739 (R2012a) 64-bit (glnxa64).

4.1. Sylvester equations in image restoration. Besides their use in control theory, Sylvester equations also appear in restoration problems for degraded images. We give a brief recapitulation of the discussions in [15, 16, 18]. Consider an image represented by a matrix $\mathbf{F} \in \mathbb{R}^{n\times m}$ with grayscale pixel values $\mathbf{F}_{ij}$ between 0 and 255. Unfortunately, often the matrix $\mathbf{F}$ is not given exactly but is perturbed by some noise or blurring process. The result is a degraded image $\mathbf{G} \in \mathbb{R}^{n\times m}$ that is obtained after an out-of-focus or atmospheric blur. One way to compute an approximately restored image $\mathbf{X} \approx \mathbf{F}$ is given by the solution to a regularized linear discrete ill-posed problem of the form

$$\min_{\mathbf{x}} \|\mathbf{H}\mathbf{x} - \mathbf{g}\|_2^2 + \lambda\|\mathbf{L}\mathbf{x}\|_2^2\,. \tag{4.1}$$

Here, $\mathbf{x} = \operatorname{vec}(\mathbf{X})$, $\mathbf{g} = \operatorname{vec}(\mathbf{G})$, $\mathbf{H}$ models the degradation process and $\mathbf{L}$ is a regularization operator with regularization parameter λ. The solution to (4.1) can be computed by solving the linear system

$$(\mathbf{H}^T\mathbf{H} + \lambda^2\mathbf{L}^T\mathbf{L})\mathbf{x} = \mathbf{H}^T\mathbf{g}.$$

While the choice of an appropriate or optimal parameter λ is a nontrivial task, we rather want to focus on efficiently solving the linear system once λ has been determined. This can, for example, be done by using the L-curve criterion or the generalized cross validation method; see [22, 29]. Following, e.g., [15], assuming certain separability properties of the blurring matrix $\mathbf{H} = \mathbf{H}_2 \otimes \mathbf{H}_1$ and the regularization operator $\mathbf{L} = \mathbf{L}_2 \otimes \mathbf{L}_1$, problem (4.1) has a special structure and can equivalently be solved by the Sylvester equation

$$(\mathbf{H}_1^T\mathbf{H}_1)\mathbf{X}(\mathbf{H}_2^T\mathbf{H}_2) + \lambda^2(\mathbf{L}_1^T\mathbf{L}_1)\mathbf{X}(\mathbf{L}_2^T\mathbf{L}_2) = \mathbf{G}. \tag{4.2}$$

In particular, we note that the matrices defining the matrix equation are symmetric positive (semi-)definite. Before we proceed, we mention typical structures of $\mathbf{H}$ and $\mathbf{L}$ that we take up

ETNA
Kent State University
http://etna.math.kent.edu

(a) Original image. (b) Blurred and noisy image.

(c) Restored image (exact), $\lambda_{\text{opt}} = 0.079$. (d) Restored image (appr.), $\lambda_{\text{opt}} = 0.127$.

Fig. 4.1: Uniform blur $(r_1 = 5)$ and atmospheric blur $(\sigma = 7, r_2 = 2)$ for $n_r = 40$.

again in the numerical examples. Again, we follow the more detailed discussions in [15, 16]. A uniform out-of-focus blur for example can be modeled by the uniform Toeplitz matrix

$$\mathbf{U}_{ij} = \begin{cases} \frac{1}{2r-1} & |i-j| \leq r, \\ 0 & \text{otherwise.} \end{cases} \tag{4.3}$$

Atmospheric blur can be realized by a Gaussian Toeplitz matrix

$$\mathbf{T}_{ij} = \begin{cases} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(i-j)^2}{2\sigma^2}\right) & |i-j| \leq r, \\ 0 & \text{otherwise.} \end{cases} \tag{4.4}$$

As in [15, 16], given an original image $\mathbf{X}$, we use out-of-focus-blur (4.3) and atmospheric blur (4.4) to construct a blurred image $\hat{\mathbf{G}}$. The final degraded image $\mathbf{G}$ is then obtained by adding Gaussian white noise $\mathbf{N}$ to $\hat{\mathbf{G}}$ such that $\frac{\|\mathbf{N}\|}{\|\hat{\mathbf{G}}\|} = 10^{-2}$.

Lothar Reichel. Due to the already mentioned dedication of this special issue, the first example is an image showing Lothar Reichel[1]. The matrix $\mathbf{X} \in \mathbb{R}^{363\times 400}$ contains grayscale

[1]The photo is taken from http://owpdb.mfo.de/detail?photo_id=3467.

ETNA
Kent State University
http://etna.math.kent.edu

(a) Original image. (b) Blurred and noisy image.

(c) Restored image (exact), $\lambda_{\text{opt}} = 1.438$. (d) Restored image (appr.), $\lambda_{\text{opt}} = 0.127$.

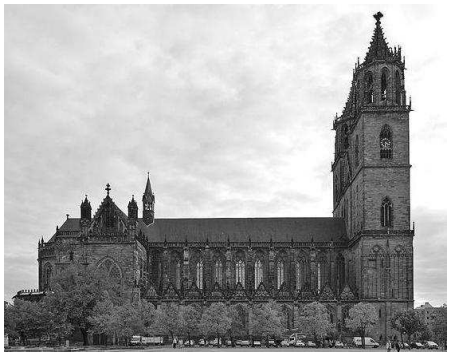
Fig. 4.2: Uniform blur $(r_1 = 6)$ and atmospheric blur $(\sigma = 12, r_2 = 6)$ for $n_r = 40$.

pixel values from the interval $[0, 255]$. The blurring matrices $\mathbf{H}_1$ and $\mathbf{H}_2$ in (4.2) are Toeplitz matrices as in (4.3) and (4.4). First, we construct $\mathbf{H}_1$ with $r_1 = 5$ and $\mathbf{H}_2$ with $\sigma = 7$ and $r_2 = 2$. We got inspired by the values chosen in [15, 16]. For the regularization operators we use discrete first-order derivatives such that

$$\mathbf{L}_1 = \begin{bmatrix} 1 & -1 & & \\ & \ddots & \ddots & \\ & & 1 & -1 \\ & & & 0 \end{bmatrix}, \quad \mathbf{L}_2 = \begin{bmatrix} 1 & & & \\ -1 & \ddots & & \\ & \ddots & 1 & \\ & & -1 & 0 \end{bmatrix}.$$

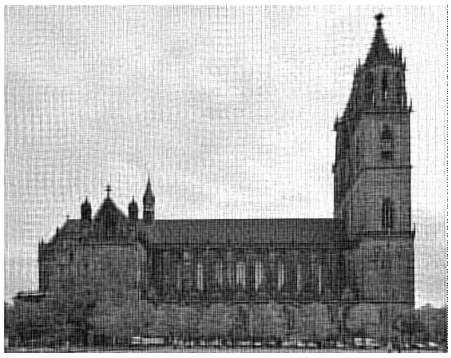
In Figure 4.1d we show the results obtained by Algorithm 1 for $n_r = 40$. We obtain a relative change less than 10^{-2} after 10 iterations. Recall that we also approximate the degraded image $\mathbf{G}$ by a low rank matrix of rank 40. We compare our result with the reconstructed image obtained by solving the Sylvester equation exactly by means of the Bartels-Stewart algorithm (4.1c). For both variants, the optimal value of the regularization parameter λ_{opt} is computed by minimization over a logarithmically spaced interval $[10^{-3}, 10]$ with 20 points. Figure 4.1 shows that the quality of the approximately reconstructed image is similar to that

ETNA
Kent State University
http://etna.math.kent.edu

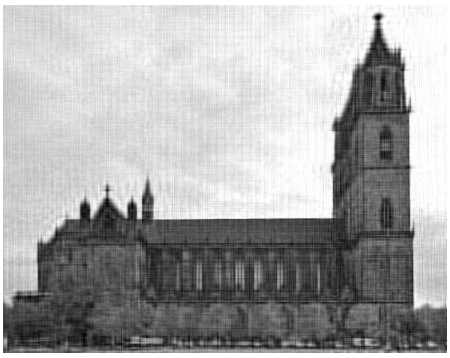


(a) Original image.

(b) Blurred and noisy image.



(c) Restored image (exact), $\lambda_{\mathrm{opt}} = 0.070$.



(d) Restored image (appr.), $\lambda_{\mathrm{opt}} = 0.298$.

Fig. 4.3: Uniform blur $(r_1 = 4)$ and atmospheric blur $(\sigma = 7, r_2 = 5)$ for $n_r = 50$.

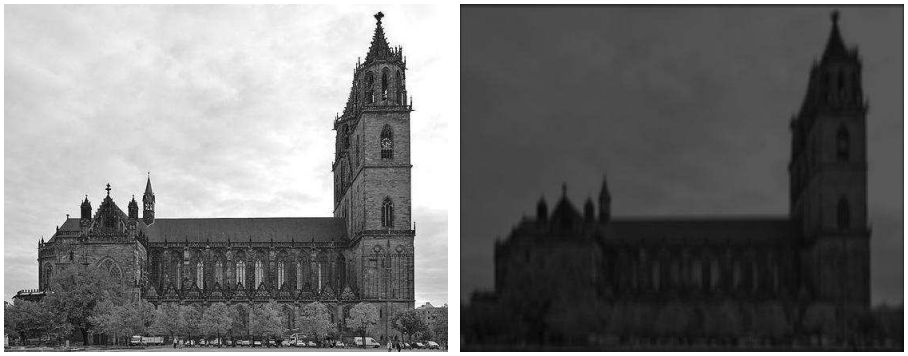
of the exactly reconstructed image. Actually, in terms of the relative spectral norm error, Algorithm 1 (0.0185) outperforms the full solution (0.1260).

Figure 4.2 shows similar results for different blurring matrices. Here, we choose $r_1 = 6$, $\sigma = 12$, and $r_2 = 6$. While the quality of the reconstructed images clearly is worse than in the first setting, Algorithm 1 obviously yields far better results than we obtain by solving the Sylvester equation explicitly. Moreover, the final (energy norm optimal) iterate from Algorithm 1 is found after 20 iteration steps.

Magdeburg cathedral. The second example is an image from the cathedral in Magdeburg, Germany[2] . The matrix $\mathbf{X}$ is of size 436×556. We choose $r_1 = 4, \sigma = 7$, and $r_2 = 5$. Since the Sylvester equation is larger than in the first example, we increase the rank of the approximation to $n_r = 50$. Figure 4.3 shows a similar comparison as in the first example. Algorithm 1 needs 19 steps before convergence is obtained. Again, the relative spectral norm error for the approximate solution (0.018) is smaller than for the exact solution (2.890). We get similar results for the parameter values $r_1 = 5, \sigma = 7$, and $r_2 = 2$. The results are shown in Figure 4.4. The number of iterations needed in Algorithm 1 is 13. Once more, note that the method used for reconstruction is probably not the most sophisticated and explains the modest quality of the approximations. Still, we point out that the reconstructed images computed by an approximate solution of the Sylvester equation in all cases perform better than

[2]The photo is taken from
http://commons.wikimedia.org/wiki/File:Magdeburger_Dom_Seitenansicht.jpg.

ETNA
Kent State University
http://etna.math.kent.edu



(a) Original image. (b) Blurred and noisy image.

(c) Restored image (exact), $\lambda_{\mathrm{opt}} = 0.336$. (d) Restored image (appr.), $\lambda_{\mathrm{opt}} = 0.055$.

Fig. 4.4: Uniform blur $(r_1 = 5)$ and atmospheric blur $(\sigma = 7, r_2 = 2)$ for $n_r = 50$.

the actual exact solution. This might be due to the badly conditioned matrices which may cause numerical perturbations when one tries to compute the full solution explicitly.

5. Conclusions. In this paper, we have studied symmetric Sylvester equations arising in dynamical control systems. The symmetric structure of the equation allows to measure errors of low-rank approximations in terms of an energy norm induced by the Sylvester operator. For a given rank n_r, we have derived first-order optimality conditions for an approximation optimal with respect to this energy norm. We have then established a connection to the $\mathcal{H}_2$-inner product of two symmetric state space systems. The corresponding first-order optimality conditions have been shown to be equivalent to the ones related to the energy norm minimization problem. The stationary points of the $\mathcal{H}_2$-inner product itself have been shown to be necessarily saddle points. An iterative interpolatory procedure trying to find these saddle points has been suggested. The two numerical examples reported and many similar experiments not described here demonstrate the applicability of the method.

REFERENCES

[1] A. C. ANTOULAS, *Approximation of Large-Scale Dynamical Systems*, SIAM, Philadelphia, 2005.
[2] A. C. ANTOULAS, C. BEATTIE, AND S. GUGERCIN, *Interpolatory model reduction of large-scale dynamical systems*, in Efficient Modeling and Control of Large-Scale Systems, J. Mohammadpour and K. Grigoriadis, eds., Springer, New York, 2010, pp. 3–58.

ETNA
Kent State University
http://etna.math.kent.edu

[3] A. C. ANTOULAS, D. C. SORENSEN, AND Y. ZHOU, *On the decay rate of Hankel singular values and related issues*, Systems Control Lett., 46 (2002), pp. 323–342.
[4] M. ATHANS, *The matrix minimum principle*, Information and Control, 11 (1967), pp. 592–606.
[5] R. H. BARTELS AND G. W. STEWART, *Algorithm 432: Solution of the matrix equation $AX + XB = C$*, Comm. ACM, 15 (1972), pp. 820–826.
[6] U. BAUR, *Low rank solution of data-sparse Sylvester equations*, Numer. Linear Algebra Appl., 15 (2008), pp. 837–851.
[7] C. BEATTIE AND S. GUGERCIN, *A trust region method for optimal H_2 model reduction*, in Proceedings of the 48th IEEE Conference on Decision and Control, IEEE Conference Proceedings, Los Alamitos, CA, 2009, pp. 5370–5375.
[8] P. BENNER, *Factorized solution of Sylvester equations with applications in control*, in Proceedings of the 16th International Symposium on Mathematical Theory of Networks and Syststems MTNS 2004, V. Blondel, P. Van Dooren, J. Willems, B. Motmans, and B. De Moor, eds., University of Leuven, Leuven, Belgium, 2004 (10 pages).
[9] P. BENNER AND T. BREITEN, *On optimality of approximate low rank solutions of large-scale matrix equations*, Systems Control Lett., 67 (2014), pp. 55–64.
[10] P. BENNER, J.-R. LI, AND T. PENZL, *Numerical solution of large Lyapunov equations, Riccati equations, and linear-quadratic control problems*, Numer. Linear Algebra Appl, 15 (2008), pp. 755–777.
[11] P. BENNER, R.-C. LI, AND N. TRUHAR, *On the ADI method for Sylvester equations*, J. Comput. Appl. Math., 233 (2009), pp. 1035–1045.
[12] P. BENNER AND E. S. QUINTANA-ORTÍ, *Solving stable generalized Lyapunov equations with the matrix sign function*, Numer. Algorithms, 20 (1999), pp. 75–100.
[13] P. BENNER AND J. SAAK, *Numerical solution of large and sparse continuous time algebraic matrix Riccati and Lyapunov equations: a state of the art survey*, GAMM-Mitt., 36 (2013), pp. 32–52.
[14] M. BOLLHÖFER AND A. EPPLER, *Structure-preserving GMRES methods for solving large Lyapunov equations*, in Progress in Industrial Mathematics at ECMI 2010, M. Günther, A. Bartel, M. Brunk, S. Schoeps, and M. Striebel, eds., Springer, Berlin, 2012, pp. 131–136.
[15] A. BOUHAMIDI AND K. JBILOU, *An iterative method for Bayesian Gauss-Markov image restoration*, Appl. Math. Model., 33 (2009), pp. 361–372.
[16] A. BOUHAMIDI, K. JBILOU, L. REICHEL, AND H. SADOK, *A generalized global Arnoldi method for ill-posed matrix equations*, J. Comput. Appl. Math., 236 (2012), pp. 2078–2089.
[17] A. BUNSE-GERSTNER, D. KUBALINSKA, G. VOSSEN, AND D. WILCZEK, *h_2-norm optimal model reduction for large-scale discrete dynamical MIMO systems*, Tech. Report 07–04, Zentrum für Technomathematik, Universität Bremen, 2007.
[18] D. CALVETTI AND L. REICHEL, *Application of ADI iterative methods to the restoration of noisy images*, SIAM J. Matrix Anal. Appl., 17 (1996), pp. 165–186.
[19] V. DRUSKIN, L. KNIZHNERMAN, AND V. SIMONCINI, *Analysis of the rational Krylov subspace and ADI methods for solving the Lyapunov equation*, SIAM J. Numer. Anal., 49 (2011), pp. 1875–1898.
[20] G. FLAGG, C. BEATTIE, AND S. GUGERCIN, *Convergence of the iterative rational Krylov algorithm*, Systems Control Lett., 61 (2012), pp. 688–691.
[21] G. FLAGG AND S. GUGERCIN, *On the ADI method for the Sylvester equation and the optimal-$\mathcal{H}_2$ points*, Appl. Numer. Math., 64 (2013), pp. 50–58.
[22] G. H. GOLUB, M. HEATH, AND G. WAHBA, *Generalized cross-validation as a method for choosing a good ridge parameter*, Technometrics, 21 (1979), pp. 215–223.
[23] G. H. GOLUB AND C. VAN LOAN, *Matrix Computations*, 3rd ed., Johns Hopkins University Press, Baltimore, 1996.
[24] L. GRASEDYCK, *Existence and computation of low Kronecker-rank approximations for large linear systems of tensor product structure*, Computing, 72 (2004), pp. 247–265.
[25] ———, *Existence of a low rank or H-matrix approximant to the solution of a Sylvester equation*, Numer. Linear Algebra Appl., 11 (2004), pp. 371–389.
[26] E. GRIMME, *Krylov Projection Methods For Model Reduction*, PhD. Thesis, Graduate College at the University of Illinois, Urbana-Champaign, 1997.
[27] S. GUGERCIN, A. ANTOULAS, AND S. BEATTIE, *$\mathcal{H}_2$ model reduction for large-scale dynamical systems*, SIAM J. Matrix Anal. Appl., 30 (2008), pp. 609–638.
[28] S. HAMMARLING, *Numerical solution of the stable, non-negative definite Lyapunov equation*, IMA J. Numer. Anal., 2 (1982), pp. 303–323.
[29] P. HANSEN, *Analysis of discrete ill-posed problems by means of the L-curve*, SIAM Rev., 34 (1992), pp. 561–580.
[30] I. M. JAIMOUKHA AND E. M. KASENALLY, *Krylov subspace methods for solving large Lyapunov equations*, SIAM J. Numer. Anal., 31 (1994), pp. 227–251.
[31] K. JBILOU AND A. J. RIQUET, *Projection methods for large Lyapunov matrix equations*, Linear Algebra Appl., 415 (2006), pp. 344–358.

ETNA
Kent State University
http://etna.math.kent.edu

[32] D. KRESSNER AND C. TOBLER, *Low-rank tensor Krylov subspace methods for parametrized linear systems*, SIAM J. Matrix Anal. Appl., 32 (2011), pp. 1288–1316.
[33] J.-R. LI AND J. WHITE, *Low rank solution of Lyapunov equations*, SIAM J. Matrix Anal. Appl., 24 (2002), pp. 260–280.
[34] A. J. MAYO AND A. C. ANTOULAS, *A framework for the solution of the generalized realization problem*, Linear Algebra Appl., 425 (2007), pp. 634–662.
[35] T. PENZL, *A cyclic low-rank Smith method for large sparse Lyapunov equations*, SIAM J. Sci. Comput., 21 (1999/2000), pp. 1401–1418.
[36] ———, *Eigenvalue decay bounds for solutions of Lyapunov equations: the symmetric case*, Systems Control Lett., 40 (2000), pp. 139–144.
[37] Y. SAAD, *Numerical solution of large Lyapunov equations*, in Signal processing, scattering and operator theory, and numerical methods MTNS-89, M. A. Kaashoek, J. H. van Schuppen, and A. C. M. Ran, eds., vol. 5 of Progr. Systems Control Theory, Birkhäuser, Boston, 1990, pp. 503–511.
[38] V. SIMONCINI, *A new iterative method for solving large-scale Lyapunov matrix equations*, SIAM J. Sci. Comput., 29 (2007), pp. 1268–1288.
[39] ———, *Computational methods for linear matrix equations*, Tech. Teport, Department of Mathematics, University of Bologna, 2013.
[40] B. VANDEREYCKEN, *Riemannian and multilevel optimization for rank-constrained matrix problems*, PhD Thesis, Department of Computer Science, Katholieke Universiteit Leuven, Leuven, Belgium, 2010.
[41] B. VANDEREYCKEN AND S. VANDEWALLE, *A Riemannian optimization approach for computing low-rank solutions of Lyapunov equations*, SIAM J. Matrix Anal. Appl., 31 (2010), pp. 2553–2579.
[42] Y. ZHOU AND D. C. SORENSEN, *Approximate implicit subspace iteration with alternating directions for LTI system model reduction*, Numer. Linear Algebra Appl., 15 (2008), pp. 873–886.