COMPUTATIONAL METHODS FOR LINEAR MATRIX EQUATIONS*

V. SIMONCINI†

Abstract. Given the matrices A, B, C, D and E of conforming dimensions, we consider the linear matrix equation $A\mathbf{X}E + D\mathbf{X}B = C$ in the unknown matrix $\mathbf{X}$. Our aim is to provide an overview of the major algorithmic developments that have taken place in the past few decades in the numerical solution of this and of related problems, which are becoming a reliable tool in the numerical formulation of advanced application models.

Key words. Sylvester equation, Lyapunov equation, Stein equation, multiple right-hand side, generalized matrix equations. Schur decomposition. Large scale computation.

AMS subject classifications. 65F10, 65F30, 15A06

Contents

*Version of March 12, 2013.

†Dipartimento di Matematica, Università di Bologna, Piazza di Porta San Donato 5, I-40127 Bologna, Italy (valeria.simoncini@unibo.it).

1. Introduction. Given the real or complex matrices A, D, E, B and C of conforming dimensions, we consider the linear matrix equation

$$A\mathbf{X}E + D\mathbf{X}B = C \tag{1.1}$$

in the unknown matrix[1] $\mathbf{X}$, and its various generalizations. If E and D are identity matrices, then (1.1) is called the Sylvester equation, as it first appearance is usually associated with the work of J.J. Sylvester [247]; if in addition $B = A^*$, where A^* is the conjugate transpose of A, then the equation is called the Lyapunov equation in honor of A. M. Lyapunov and his early contributions to the stability problem of motion; see [14],[186] and the whole issue of the same journal. We shall mainly consider the generic case, thus assuming that the all involved matrices are nonzero.

Under certain conditions on the coefficient matrices, (1.1) has a unique solution, with available elegant and explicit closed forms. These are usually inappropriate as computational devices, either because they involve estimations of integrals, or because they can be polluted with numerical instabilities of various sorts. Nonetheless, closed forms and other properties of the solution matrix have strongly influenced the computational strategies that have lead to most algorithms used today for numerically solving (1.1), in the case of small or large dimensions of the coefficient matrix. Due to the availability of robust and reliable core algorithms, (1.1) now arises in an increasingly larger number of scientific computations, from statistics to dynamical systems analysis - with a major role in control applications, and also as a workhorse of more intensive computational methods. In section 2 we will briefly review this broad range of numerical and application problems.

Our aim is to provide an overview of the major algorithmic developments that have taken place in the past few decades in the numerical solution of (1.1) and of related problems, both in the small and large scale cases. A distinctive feature in the large-scale setting is that although the coefficient matrices may be sparse, the solution matrix is usually dense and thus impossible to store in memory. Therefore, ad-hoc strategies need to be devised to approximate the exact solution in an affordable manner.

Functions related to the solution matrix $\mathbf{X}$ such as the spectrum, the trace and the determinant, also have an important role in stability analysis and other applications. Although we shall not discuss these problems in detail, we shall occasionally point to relevant results and appropriate references.

The Lyapunov equation has received a lot of attention, mainly thanks to its prominent role in control. In particular, many authors have focused on numerical strategies associated specifically to this equation, making the corresponding literature particularly rich. As a consequence, the Sylvester and Lyapunov equations have somehow evolved differently. For these reasons, and to account for the literature in a homogeneous way, we shall first discuss numerical strategies for the Sylvester equation, and then treat in detail the Lyapunov problem, emphasizing the overlapping solution methods. It should be noticed, however, that all numerical solution strategies available for the Sylvester equation can be used for the Lyapunov case, and that many of the Lyapunov solvers can be naturally adapted to the general case. For A and B of size up to a few hundreds, the Schur-decomposition based algorithm by Bartels and Stewart ([15]) has since its appearance become the main numerical solution tool. In the large scale case, various directions have been taken, and a selection of effective

[1] Here and in the following we shall use bold face letters to denote the unknown solution matrices.

algorithms is available, from projection methods to sparse format iterations, with no clear winner for all settings. The distinction between small, moderate and large size is clearly architecture dependent. In the following we shall refer to "small" and medium problem size when the coefficient matrices have dimensions of a few thousands at most. These linear equations can be solved with decomposition-based methods on laptops with moderate computational efforts. The target for current large-scale research are matrices of dimensions $\mathcal{O}(10^6)$ or larger, with a variety of sparsity patterns.

Throughout the paper we shall assume that E, D are either the identity, or are nonsingular matrices. Singular E, D have great relevance in control applications associated with differential-algebraic equations and descriptor systems but require a specialized treatment, which can be found, for instance, in [166].

Equation (1.1) is a particular case of the linear matrix equation

$$A_1\mathbf{X}B_1 + A_2\mathbf{X}B_2 + \dots A_k\mathbf{X}B_k = C, \tag{1.2}$$

with A_i, B_i, $i = 1, \dots, k$ square matrices, and C of dimension $n \times m$. While up to 15-20 years ago this multi-term equation could be rightly considered of mainly theoretical interest, the dramatic recent developments associated with problems stemming from applications with a dominant stochastic component have brought generalized linear matrix equations to play a fundamental role; see also section 2 and section 7.2.

Equation (1.2) is very difficult to analyze in its full generality, and necessary and sufficient conditions for the existence and uniqueness of the solution $\mathbf{X}$ explicitly based on $\{A_i\}, \{B_i\}$, are hard to get, except for some very special cases [170],[159]. The numerical solution of (1.2) has been largely overlooked, and this is particularly so in the large-scale setting, namely when either or both A_i and B_i have large dimensions. Moreover, while from a theoretical view point the importance of taking into account the structure of the problem has been acknowledged [159], this is not so for computational strategies, especially with large scale problems. The main algorithmic device for (1.2) consists in transforming the matrix above into a vector form by means of the Kronecker product (defined below). Such basic approach was abandoned for (1.1) as a core method, since algorithms with a complexity of a modest power of the coefficient matrices dimension are now available. The efficient numerical solution to (1.2) thus represents the next frontier for linear matrix equations, so as to assist quickly developing application problems.

Various forms of generalizations have also been tackled, as they are more and more often encountered in applications. This is the case, for instance, for bilinear equations (in two unknown matrices), and for *systems* of bilinear equations. These represent an open computational challenge, and their efficient numerical solution would provide a great advantage for emerging mathematical models.

A very common situation arises when $B = 0$ and C is tall in (1.1), so that the matrix equation reduces to a standard linear system with multiple right-hand sides, the columns of C. This is an important problem, and it is often encountered in applications; a significant body of literature is available, with a large number of contributions in the past twenty years. Since the most popular procedures for the solution of $A\mathbf{X} = C$ are usually derived from the single right-hand side case, we shall not discuss them here, as the topic surely deserves a dedicated treatment; instead, we refer to [220] and to the recent list of references [114].

After a brief account of the numerous application problems where linear matrix equations arise, we shall recall the main properties of these equations, together with possible explicit forms for their solution matrix. The rest of this paper describes

many approaches that have been proposed in the recent literature: we first treat the Sylvester equation, when A and B are small, when one of the two is large, and when both are large. Indeed, rather different approaches can be employed depending on the size of the two matrices. We shall then focus on the Lyapunov equation: due to its relevance in control, many developments have specifically focused on this equation, therefore the problem deserves an ad-hoc treatment. The small-scale problem is computationally well understood, whereas the large-scale case has seen quite significant developments in the past ten years. Later sections report on the computational devices associated with the numerical solution of various generalizations of (1.1), which have been taking place in the past few years.

2. Applications. Matrix equations are ubiquitous in signal processing, control and system theory; see, e.g., [4],[258],[91],[70],[23],[28],[224],[63] and references therein. Most time-dependent models may be represented as linear or non-linear dynamical systems, accounting for the prediction, simulation and control of real world phenomena. The numerical solution of matrix equations and the relevance of its role within engineering applications justify the great effort put into this problem by the scientific community, both from the mathematical and from the applied directions. Special issues of journals and multi-contribution books are often devoted to advances in this and related areas, attesting the continuous pressure for up-front numerical devices, that can take into account the properties of the problem, such as structure, size and operational characterizations.

Linear matrix equations have an important role in the stability analysis of linear dynamical systems, and take also part in the theoretical developments of non-linear ones. Consider the following continuous-time linear system[2]

$$\dot{x} = Ax + B_1 u \tag{2.1}$$

$$y = B_2^\top x \tag{2.2}$$

where x is the model state, u is the input and y is the output, and the matrices A, B_1 and B_2 are time-invariant. Assuming A is stable, that is its eigenvalues have negative real part, then the solutions $\mathbf{P}$ and $\mathbf{Q}$ to the following Lyapunov equations

$$A\mathbf{P} + \mathbf{P}A^\top + B_1 B_1^\top = 0, \quad A^\top \mathbf{Q} + \mathbf{Q}A + B_2 B_2^\top = 0,$$

are called the controllability and observability Gramians, respectively, and they are used, for instance, to control the energy of the system [4, sec.4.3.1]. Under certain additional hypotheses it may be shown that the symmetric matrices $\mathbf{P}$ and $\mathbf{Q}$ are positive definite. These two latter matrices are key when one is interested in reducing the original system into one of much smaller dimension, while essentially preserving the main dynamical system properties. Indeed, one of the goals of *balanced reduction*, is to determine an appropriate representation basis for the system such that the Gramians are equal and diagonal [193], so that the reduction of that basis will maintain this property of the Gramians. The diagonal Gramians then contain information on the output error induced by the reduced model.

Alternatively, if B_1 and B_2 have the same number of columns, one can solve the following Sylvester equation,

$$A\mathbf{W} + \mathbf{W}A + B_1 B_2^\top = 0,$$

[2] In the control literature, B_1, B_2 are usually denoted by B and $C^\top$, respectively; we opted for a slightly different notation because here B and C have a different meaning.

thus obtaining the *cross-Gramian* $\mathbf{W}$ [88], which contains information on controllability and observability of the system. For B_1, B_2 having a single column, or for A symmetric and B_1, B_2 such that $B_2^\top (zI - A)^{-1} B_1$ is symmetric, it is possible to show that $\mathbf{W}^2 = \mathbf{PQ}$, so that the eigenvalues of $\mathbf{W}$ coincide with the square root of the eigenvalues of $\mathbf{PQ}$ [89],[242]. In general, the latter are called the Hankel singular values of the system, and they are invariant under state space transformations; we refer to [4] for a detailed discussion of these quantities and their role in model order reduction. A different Sylvester equation was used in [92] to derive a numerical algorithm that couples the two Gramians $\mathbf{P}$ and $\mathbf{Q}$. Similar results can be stated for the case of the discrete-time time-invariant linear systems

$$\begin{aligned} x(t+1) &= Ax(t) + B_1 u(t) \\ y(t) &= B_2^\top x(t), \end{aligned}$$

which are associated, for instance, with the discrete-time Lyapunov equation

$$A\mathbf{X}A^\top - \mathbf{X} + B_1 B_1^\top = 0.$$

Linear matrix equations are also used in control as a technical tool for solving other problems, see, e.g., [22],[286],[91],[172],[198], and for the reduction of nonlinear models; see, e.g., [265],[160],[294],[62] and references therein.

The Sylvester equation often occurs in linear and generalized eigenvalue problems for the computation of invariant subspaces by means of the Riccati equation [246], [235], [73]. In fact, the algebraic Riccati equation itself provides a formidable setting for linear matrix equations: this quadratic equation is sometimes dealt with by solving a sequence of linear Sylvester equations with possibly varying known term and coefficient matrices [43]. With the aim of controlling resonance modes in vibrating structures, Sylvester equations also arise in solving quadratic eigenvalue assignment problems, see, e.g., [50]. Large eigenvalue problems are also a key step in the detection of a Hopf bifurcation in large-scale dynamical systems that depend on some physical parameters. However, it is possible to compute these parameters without actually computing the relevant eigenvalues. In [189], it was shown that this can be performed by means of a *matrix* inverse iteration procedure, which involves approximately solving a sequence of large-scale Lyapunov equations. Lyapunov equations are a theoretical and computational tool also in hydrodynamic stability theory of time-dependent problems, which is emerging as an attractive alternative to classical modal spectral analysis, in the quantitative description of short-term disturbance behaviors [225].

Different application areas have emerged that can take advantage of an efficient solution of linear matrix equations. Problems associated with image processing seem to provide a rich source. For instance, Sylvester equations can be used to formulate the problem of restoration of images affected by noise [55]: an optimization model is formulated describing the image-object correspondence, and noise is included. The minimum of the cost functional can be explicitly obtained, and the estimated image can be reconstructed as the solution of a linear matrix equation. A similar optimization model can be used in adaptive optics, a technology developed for compensation of aberrations in optical systems or due to atmospheric turbulence, mainly used in high quality astronomical observations and measurements [214]. Within the image application, the problem of estimating a 3D object's pose obtained from 2D image sequences can be stated as a constrained optimization problem [60], [59]. This leads

to the solution of a sequence of small Sylvester equations. In fact, depending on the number of poses, the occurring linear matrix equations have more than two terms, and can be formulated as in (1.2); see [59].

The Sylvester equation was highlighted as a *model problem* in the solution of elliptic boundary value problems governed by the two-dimensional differential operator

$$\mathcal{L}(u) = -\nabla \cdot (\kappa \nabla u)$$

by Ellner and Wachspress [83]: they devised a matrix algorithmic version of the (differential) ADI algorithm by Peaceman and Rachford, and this became the founder of ADI-type methods for linear matrix equations. Wachspress showed that the constant coefficient second order differential equation can be used as preconditioner for the original operator, and that the application of the preconditioner amounts to solving a Lyapunov equation [270]. Sylvester equations can also be used in the implementation of implicit Runge-Kutta integration formulae and block multistep formulae for the numerical solution of ordinary differential equations [86]. Discrete-time Sylvester and Lyapunov equations (cf. section 6) also arise for instance in statistics and probability ([155],[154],[153],[10]), and as a building block of discrete-time algebraic Riccati equation [43]. A large list of references on application problems where the Lyapunov equation plays an important role is available in the last chapter of [91].

3. Notation and preliminary definitions. Unless stated otherwise, throughout the paper we shall assume that the coefficient matrices are real. Moreover, $\mathrm{spec}(A)$ denotes the set of eigenvalues of A, and $A^\top$, A^* denote that transpose and conjugate transpose of A, respectively. A matrix A is *stable* if all its eigenvalues have negative real part, and *passive* if for all unit 2-norm complex vectors x, the quantity x^*Ax has negative real part; the term "negative definite" is sometimes also used instead of "passive". The notation $A > 0$ ($A \geq 0$) states that A is a symmetric and positive definite (semi-definite) matrix.

The vector e_i denotes the ith column of the identity matrix, whose dimension will be clear from the context; I_n denotes the identity matrix of size n, and the subscript will be omitted when clear from the context. Throughout, given $x \in \mathbb{C}^n$, $\|x\|$ denotes the 2-norm of x, $\|A\|$ or $\|A\|_2$ denotes the matrix norm induced by the vector 2-norm, while $\|A\|_F$ denotes the Frobenius norm of $A = (a_{i,j})_{i=1,\dots,n,j=1,\dots m}$, that is $\|A\|_F^2 = \sum_{i,j} |a_{i,j}|^2$. The notation $[A; B]$ will be often used to express the matrix obtained by stacking the matrix B below the matrix A, both having conforming dimensions.

For given matrices $A \in \mathbb{C}^{n_A \times m_A}$, $A = (a_{ij})_{i=1,\dots,n_A, j=1,\dots,m_A}$ and $B \in \mathbb{C}^{n_B \times m_B}$, the Kronecker product is defined as

$$A \otimes B = \begin{bmatrix} a_{11}B & a_{12}B & \cdots & a_{1m_A}B \\ a_{21}B & a_{22}B & \cdots & a_{2m_A}B \\ \vdots & & & \vdots \\ a_{n_A1}B & a_{n_A2}B & \cdots & a_{n_Am_A}B \end{bmatrix} \in \mathbb{C}^{n_An_B \times m_Am_B};$$

the vec operator stacks the columns of a matrix $X = [x_1, \dots, x_m] \in \mathbb{C}^{n \times m}$ one after the other as

$$\mathrm{vec}(X) = \begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix} \in \mathbb{C}^{nm \times 1}.$$

We summarize some well known properties of the Kronecker product in the next lemma.

LEMMA 3.1. (cf., e.g., [132]) *Some properties:*

(i) $vec(AXB) = (B^\top \otimes A)vec(X)$ *(note the complex transposition for* B*);*

(ii) If $A \in \mathbb{R}^{n\times n}$ *and* $B \in \mathbb{R}^{m\times m}$*, and* $\lambda_A \in \mathrm{spec}(A)$*,* $\lambda_B \in \mathrm{spec}(B)$*, then* $\lambda_A\lambda_B \in \mathrm{spec}(A \otimes B)$*; (and every eigenvalue of* $A \otimes B$ *is the product of eigenvalues of* A *and* B*)*

(iii) Under the hypotheses of (ii), $\lambda_A + \lambda_B \in \mathrm{spec}(I_m \otimes A + B \otimes I_n)$*; (and every eigenvalue of* $I_m \otimes A + B \otimes I_n$ *is the sum of eigenvalues of* A *and* B*)*

4. Continuous-time Sylvester equation. The continuous-time Sylvester equation is possibly the most broadly employed linear matrix equation, and is given as

$$A\mathbf{X} + \mathbf{X}B = C, \tag{4.1}$$

with $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{m\times m}$ and $C \in \mathbb{R}^{n\times m}$. In general, the dimensions of A and B may be orders of magnitude different, and this fact is key in selecting the most appropriate numerical solution strategy.

A general result on the consistency of the Sylvester equation was given by Roth in 1952 in [217], and it reads: *the equation (4.1) admits a solution if and only if the matrices*

$$\begin{bmatrix} A & -C \\ 0 & -B \end{bmatrix} \quad \textit{and} \quad \begin{bmatrix} A & 0 \\ 0 & -B \end{bmatrix} \tag{4.2}$$

are similar.

Using the Kronecker product, this matrix equation can be rewritten as the following standard (vector) linear system

$$\mathcal{A}\mathbf{x} = c, \qquad \text{with} \quad \begin{array}{l} \mathcal{A} = I_m \otimes A + B^\top \otimes I_n \\ \mathbf{x} = \mathrm{vec}(\mathbf{X}), \quad c = \mathrm{vec}(C), \end{array} \tag{4.3}$$

from which we can deduce that the system admits a solution for any c and this is unique, if and only if the matrix $\mathcal{A}$ is nonsingular. Taking into account Lemma 3.1(iii), this is equivalent to requiring that $\mathrm{spec}(A) \cap \mathrm{spec}(-B) = \emptyset$ (cf., e.g., [132, Th. 4.4.6]). In the following we shall thus always assume that this latter condition is satisfied, so that the solution to (4.1) exists and is unique; we refer to standard matrix analysis books for the case when this spectral condition is not satisfied (cf., e.g., [132], [169]). The homogeneous case, namely when $C = 0$, can be handled correspondingly: the matrix equation has only the trivial solution $\mathbf{X} = 0$ if and only if $\mathrm{spec}(A) \cap \mathrm{spec}(-B) = \emptyset$ [98, sec.17.8].

The solution $\mathbf{X}$ of (4.1) may be written in closed form in a number of different ways (see, e.g., [170], [72], [132], [141]):

(*a*) *Integral of resolvents.* The following representation, due to Krein, exploits spectral theory arguments:

$$\mathbf{X} = -\frac{1}{4\pi^2}\int_{\Gamma_1}\int_{\Gamma_2} \frac{(\lambda I_n - A)^{-1}C(\mu I_m - B)^{-1}}{\lambda + \mu} d\mu d\lambda, \tag{4.4}$$

where Γ_1, Γ_2 are contours containing and sufficiently close to, the spectra of A and B, respectively.

(*b*) *Integral of exponentials.* This representation is tightly connected to the previous one,

$$\mathbf{X} = -\int_0^\infty e^{At} C e^{Bt} dt, \tag{4.5}$$

where e^{Ht} is the matrix exponential of Ht. Here the spectra of A and B are supposed to be separated by a vertical line.

(*c*) *Finite power sum.* Let $C = C_A C_B^\top$, and let m, k be the degrees of the minimal polynomials of A with respect to C_A and of B with respect to C_B, respectively. Then

$$\mathbf{X} = \sum_{i=0}^{m-1}\sum_{j=0}^{k-1} \gamma_{ij} A^i C B^j = [C_A, AC_A, \ldots, A^{m-1}C_A](\gamma \otimes I) \begin{bmatrix} C_B^\top \\ C_B^\top B \\ \vdots \\ C_B^\top B^{k-1} \end{bmatrix}$$

where γ is the solution of the Sylvester equation with coefficient matrices the companion matrices of A and B, and right-hand side the matrix $E_1 E_1^\top$, where E_1 contains the first few columns of the identity matrix, and it has the same number of columns as C_A and C_B; see also [167].

(*d*) *Similarity transformations.* Strictly related to (c), in addition this form assumes that A and B can be diagonalized, $U^{-1}AU = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ and $V^{-1}BV = \mathrm{diag}(\mu_1, \ldots, \mu_n)$. Let $\widetilde{C} = U^{-1}CV$. Then

$$\mathbf{X} = U\widetilde{X}V^{-1}, \quad \text{with} \quad \widetilde{\mathbf{x}}_{ij} = \frac{\widetilde{c}_{ij}}{\lambda_i + \mu_j}.$$

Other representations can be found in [170], and for more general equations of the same type, in [276]. We also mention that the columns of $[\mathbf{X}; I]$ span an invariant subspace for the left matrix in (4.2), that is

$$\begin{bmatrix} A & -C \\ 0 & -B \end{bmatrix} \begin{bmatrix} \mathbf{X} \\ I \end{bmatrix} = \begin{bmatrix} \mathbf{X} \\ I \end{bmatrix} S \tag{4.6}$$

where it holds that $S = -B$. This viewpoint has allowed the derivation of several properties of the matrix equation, from stability to algorithmic devices.

In [72] the closed form in (c) is used to derive results on the solution rank. In there, the controllability (resp. observability) of the pair (A, C_A) (resp. $(B^\top, C_B)$) plays a crucial role[3]. Results on the nonsingularity of the solution based on the same conditions are also contained in [121]. For more general equations, corresponding conditions can be found, e.g., in [275].

Early computational methods relied on one of the analytic expressions above; see, e.g., [141], and the account on early methods in [91]. However, these closed forms are no longer used to numerically solve the Sylvester equation, as they are clearly inefficient - and possibly unstable - even for small n, m. On the other hand, they have been used as motivation to several successful methods and they represent an important starting point for theoretical investigations of numerical approaches.

[3] A pair (M, C) is controllable if the matrix $[C, MC, \ldots, M^{n-1}C]$ has full row rank n, equal to the row dimension of M; $(M, C^\top)$ is observable if $(M^\top, C)$ is controllable.

4.1. Stability and sensitivity issues of the Sylvester equation. In this section we provide a brief account of the most pressing problems encountered when dealing with solutions to the Sylvester equation. The topic is broad, and it also involves the solution of related matrix equations; we refer to the thorough treatment in [159] for a full account on the perturbation theory of this and other important equations in control.

The sensitivity to perturbations of the solution $\mathbf{X}$ to (4.1) is inversely proportional to the separation between A and $-B$, where the separation function of two matrices A_1 and A_2 is defined as

$$\mathrm{sep}_p(A_1, A_2) = \min_{\|P\|_p=1} \|A_1 P - P A_2\|_p,$$

with $p = 2, F$ [245][262]. This can be viewed by using the theory developed by Stewart in [245], where the columns of $[\mathbf{X}; I]$ are a basis for an invariant subspace for the first block matrix in (4.6). We refer, e.g., to [99, section 7.6.3] where the role of $\|X\|$ in the conditioning of the associated eigenvalues is emphasized. More specifically, it holds that

$$\|\mathbf{X}\| < 2\frac{\|C\|}{\mathrm{sep}_2(A, -B)}.$$

For non-normal matrices, the bound above shows that a good spectral distance between A and $-B$ might not be sufficient to limit the size of $\|\mathbf{X}\|$, since $\mathrm{sep}_2(A, -B)$ can be much smaller than the distance between the spectra of A and $-B$. The function sep is thus viewed as a condition number for the following Sylvester operator

$$\mathcal{S} : \mathbb{R}^{n\times m} \to \mathbb{R}^{n\times m}, \quad \mathcal{S}(X) = AX + XB; \tag{4.7}$$

numerical estimates for the sep function can be obtained by carefully adapting classical strategies [51]. The discussion above also shows that for small scale equations, algorithms that rely on orthogonal reduction should be preferred in terms of numerical stability. Methods that rely on more general similarity transformations may transfer the ill-conditioning of the transformation matrix onto large errors in the obtained solution [245, Th.4.8].

A major difference between matrix equations and standard linear systems lies in their stability properties. In particular, a small Sylvester equation residual does not necessarily imply a small backward error [124, sec.15.2]. Define the backward error for an approximation $\mathbf{X}$ as

$$\begin{aligned}\eta(\mathbf{X}) :=& \min\{\varepsilon : (A + \Delta A)\mathbf{X} + \mathbf{X}(B + \Delta B) = C + \Delta C,\\ & \|\Delta A\|_F \le \varepsilon\|A\|_F, \|\Delta B\|_F \le \varepsilon\|B\|_F, \|\Delta C\|_F \le \varepsilon\|C\|_F\},\end{aligned}$$

and the residual as $R = C - (A\mathbf{X} + \mathbf{X}B)$. Then ([124])

$$\eta(\mathbf{X}) \le \mu\frac{\|R\|_F}{(\|A\|_F + \|B\|_F)\|\mathbf{X}\|_F + \|C\|_F}, \tag{4.8}$$

where μ is an amplification factor depending on the data norms and on the singular values of $\mathbf{X}$. For instance, for $n = m$ this factor has the expression

$$\mu = \frac{(\|A\|_F + \|B\|_F)\|\mathbf{X}\|_F + \|C\|_F}{((\|A\|_F^2 + \|B\|_F^2)\sigma_{\min}(\mathbf{X})^2 + \|C\|_F^2)^{\frac{1}{2}}},$$

making the dependence on the norm and ill-conditioning of $\mathbf{X}$ more apparent. A more complex situation occurs for $n \neq m$; we refer to [124, sec.15.2] for more details. We also mention that in [254] bounds for the norm of the solution $\mathbf{X}$ and of its perturbation are obtained, that emphasize the influence of the possibly low-rank right-hand side on the sensitivity of the solution itself.

The distribution of the singular values of $\mathbf{X}$ plays a crucial role in the stability analysis of dynamical systems. Remarkably, such distribution is also extremely relevant in the development and convergence analysis of iterative solution methods. Indeed, it is expected that a Sylvester equation having solution with exponentially decaying singular values will be well approximated by a low rank matrix. We recall here the result described by Sabino in [221, Th.2.1.1], and we refer to Sabino's PhD thesis for further discussion related to this bound. Here K and K' are the complete elliptic integrals of the first kind[4] [1]. Additional considerations are postponed to the Lyapunov equation case in section 5.

THEOREM 4.1. *Let A and B be stable and symmetric, with spectra contained in $[a,b]$ and $[c,d]$, respectively. Define $\eta = 2(b-a)(d-c)/((a+c)(b+d))$. Assume C is of rank p. Then the singular values $\sigma_1 \geq \ldots \geq \sigma_{\min\{m,n\}}$ of the solution $\mathbf{X}$ to (4.1) satisfy*

$$\frac{\sigma_{pr+1}}{\sigma_1} \leq \left(\frac{1-\sqrt{k_r'}}{1+\sqrt{k_r'}} \right)^2, \quad 1 \leq pr < n,$$

where $k_r' = 1/(1+\eta+\sqrt{\eta(\eta+2)})$ is the complementary elliptic modulus corresponding to the nome q^r, $q := \exp(-\pi K'/K)$.

A more accessible and practical estimate for $B = A$ and small condition number $\kappa(A)$ may be obtained as ([221])

$$\frac{\sigma_{pr+1}}{\sigma_1} \lesssim 4\exp(-\pi^2 r/\log(4\kappa(A))). \tag{4.9}$$

Easy to use variants of (4.9) are compared in [221] with earlier estimates in [204]. Results for A and B nonsymmetric are few, and mainly experimental; non-normality may strongly influence the solution rank, so that results will significantly depart from the above bound. This area of research is essentially open.

From a numerical analysis viewpoint, we notice that the main ingredients from rational approximation used for results of the type above are the same as those obtained for rational space projections and ADI-type iterations (see section 4.4.1 and section 4.4.2, respectively), which also rely on minimax rational approximations; in fact, the result above is intimately related to similar estimates by Ellner and Wachspress in [83], [84].

4.2. Sylvester equation. Small scale computation. A robust and efficient method for numerically solving Sylvester equations of small and moderate size was introduced in 1972 by Bartels and Stewart [15], and with some modifications is still the state-of-the-art; in section 8 we give an account of current software, highly relying on this method. The idea is to compute the Schur decomposition of the two coefficient matrices and then transform the given equation into an equivalent one that uses the quasi-lower/upper structure of the Schur matrices. The last equation can then be

[4] They are defined as $K = K(k) = \int_0^1 [(1-t^2)(1-kt^2)]^{-1/2} dt$ and $K' = K(1-k)$, with k being the modulus, $k = \sqrt{1-(k')^2}$, while the complementary elliptic modulus k' is given.

explicitly solved element by element. More precisely, the algorithm performs the following steps (see, e.g., [99],[229]):

1. Compute the Schur forms: $A^* = URU^*$, $B = VSV^*$ with R, S upper triangular;
2. Solve $R^*\mathbf{Y} + \mathbf{Y}S = U^*CV$ for $\mathbf{Y}$;
3. Compute $\mathbf{X} = U\mathbf{Y}V^*$.

The Schur forms in the first step are obtained by the QR algorithm [99], while the third step is a simple product. It remains to explain how to solve the new structured Sylvester equation in the second step. Assuming complex arithmetic is used throughout, R^* is lower triangular and S is upper triangular, so that, for instance, the (1,1) element of $\mathbf{Y}$ can be readily obtained. From there the next elements of the first row in $\mathbf{Y}$ can also be obtained sequentially. A similar reasoning can be used for the subsequent rows. In the case of real A and B, the *real* Schur form may be exploited, where R and S are now quasi-triangular, that is the diagonals have 2×2 and 1×1 blocks, corresponding to complex and real eigenvalues, respectively. The process relies on the equivalence between a 2×2 Sylvester equation and the associated Kronecker form [229]. The same sequential process as in the complex case can be employed to compute the elements of $\mathbf{Y}$, as long as the diagonal blocks can be made conforming [15],[99]. The method just outlined is one of the workhorses of the software package SLICOT [257],[238],[35]. The leading computational cost is given by the Schur forms in step one, which for real matrices are nowadays performed in real arithmetic. Explicitly writing the Schur form costs at least $10n^3$ for a matrix of size n [99]; to limit costs, the SLICOT developers suggest to employ the Bartels-Stewart algorithm only if either A or B is already in Schur or upper Hessenberg form [238]. For general matrices A and B, the method proposed by Golub, Nash and Van Loan in 1979 ([100]) can be considerably faster, especially if either m or n is significantly smaller than the other. This latter method replaces the Schur decomposition of the larger matrix, say, B, with the Hessenberg decomposition of the same matrix whose computational cost is $5/3m^3$, which should be compared with $10m^3$ of the Schur form [100]. We refer to [229, sec.2.3.1] for a more detailed comparisons on the computational costs. In fact, these flop counts may be too pessimistic and thus possibly misleading: CPU times are in practice more similar than predicted from the flop count. In [241], a variant of the Bartels-Stewart algorithm is proposed: the forward-backward substitution in step 2 is performed by a column-wise block scheme, which seems to be better suited for modern computer architectures than the original complex version; we also refer to [146],[147] for even more effective implementations.

Iterative solution strategies for small size matrices have also been proposed: given an initial guess $\mathbf{X}_0$, they determine a sequence of matrices $\mathbf{X}_1, \ldots, \mathbf{X}_k, \ldots$ that converge to $\mathbf{X}$. These are related to a basic Newton iteration for approximating the matrix sign function; see, e.g., [22]. In section 5.2.3 we will give more details in relation with the Lyapunov equation, although the procedure can be used for stable Sylvester equations as well [39]. These approaches are easier to parallelized than QR based methods. For instance, in [39] they provide high efficiency and scalability on clusters of machine processors.

Recently, an Hermitian and skew-Hermitian splitting strategy was used in [9]; the idea could be adapted to solve large scale problems, however important computational issues such as low rank recurrences, parameter selection and residual computation still need to be addressed. The connection with classical ADI methods is also unexplored.

Finally, a numerical method based on a closed form solution was derived in [210], which used the characteristic polynomials of A and B, though the stability properties of this approach have not been analyzed. We recall that the solution can be given in closed form by using the *minimal polynomials* of A and B, if C is low rank; cf. section 4.1. For stability reasons, generic numerical procedures based on these closed forms should be avoided.

To conclude, a special mention deserves the Sylvester equation with $B = -A$, yielding the so-called *displacement equation*

$$A\mathbf{X} - \mathbf{X}A = C, \tag{4.10}$$

which measures how far A and $\mathbf{X}$ are from commuting; see [44],[97] for typical applications in the context of structured matrices such as Cauchy-like and Toeplitz matrices.

4.3. Sylvester equation. Large A and small B. When either n or m is large, reduction may require a prohibitive amount of space, due to the full decomposition of the corresponding large matrix. Selecting the most appropriate solver still depends on whether the smaller matrix has very small dimension. Different approaches can then be used when decomposing the small matrix is feasible[5]. To fix ideas, and without loss of generality, we shall assume that B is small and A is large, so that $m \ll n$.

In this section we thus consider that the equation can be visualized as:

$$\begin{bmatrix} \\ A \\ \\ \end{bmatrix} \begin{bmatrix} \\ \mathbf{X} \\ \\ \end{bmatrix} + \begin{bmatrix} \\ \mathbf{X} \\ \\ \end{bmatrix} \begin{bmatrix} B \end{bmatrix} = \begin{bmatrix} \\ C \\ \\ \end{bmatrix}, \tag{4.11}$$

so that only the large dimension of A provides new challenges with respect to section 4.2. This setting arises for instance in the solution of eigenvalue problems [274, sec.2.4, sec.6.6] and in (separable) boundary value problems [268],[272],[42]. We readily notice that for very small m, the transformation with the Kronecker product (4.3) may be appealing, since the dimension of the linear system may be just a few (m) times that of A. However, projection methods acting on the original matrix equation turn out to be extremely effective in this case, possibly justifying the fewer attempts to pursue such Kronecker formulation. We next describe some of the standard approaches currently employed.

Assume that B can be spectrally decomposed cheaply and stably. Then by writing $B = WSW^{-1}$ with $S = \mathrm{diag}(s_1, \ldots, s_m)$, we obtain

$$A\widehat{\mathbf{X}} + \widehat{\mathbf{X}}S = \widehat{C}, \quad \widehat{\mathbf{X}} = \mathbf{X}W, \quad \widehat{C} = CW. \tag{4.12}$$

For B symmetric, $W^{-1} = W^\top$. Each column of $\widehat{\mathbf{X}}$ can be obtained by solving a shifted linear system $(A + s_i I)(\widehat{\mathbf{X}})_i = (\widehat{C})_i$, where $(\widehat{\mathbf{X}})_i$ denotes the ith column of $\widehat{\mathbf{X}}$. The main steps are as follows:

1. Compute the decomposition $B = WSW^{-1}$
2. Set $\widehat{C} = CW$
3. For each i, solve $(A + s_i I)(\widehat{\mathbf{X}})_i = (\widehat{C})_i$
4. Compute $\mathbf{X} = \widehat{\mathbf{X}}W^{-1}$

[5] Feasibility is machine architecture dependent; nonetheless, a matrix of dimension much less than one thousand should be considered small.

The shifted systems in step 3 can be solved simultaneously by using standard solvers for algebraic linear systems, either direct or iterative; see, e.g., [220], [236] and their references. We also note that step 3 is "embarrassingly parallel" in case different systems can be distributed on a multiprocessor machine.

If the eigendecomposition of B is not appealing, then one can resort to a (complex) Schur decomposition $B = QR_BQ^*$, giving $A\mathbf{X}Q + \mathbf{X}QR_B = CQ$. Since R_B is upper triangular, these systems can still be solved by using the shifted form, but this time in sequence: letting r_{ij} be the (i, j) entry of R_B and $\widehat{C} = CQ$, we have

$$\text{for } i = 1, \dots, m, \quad (A + r_{ii}I)(\widehat{\mathbf{X}})_i = (\widehat{C})_i - \sum_{k=1}^{i-1} r_{ki}(\widehat{\mathbf{X}})_k, \quad \widehat{\mathbf{X}} = \mathbf{X}Q; \tag{4.13}$$

such approach has been used in different contexts, see, e.g., [111],[243],[34], where the considered Sylvester equation is occasionally called *sparse-dense* equation.

For moderate n, the use of direct methods in (4.12) and (4.13) may entail the use of complex arithmetic if the shifts (eigenvalues) are complex, significantly increasing the computational cost; the alternative of solving two real systems also lead to higher computational costs. In addition, in case the use of sparse direct methods appears to be competitive, it should be noticed that only the sparsity analysis step can be done once for all, whereas the actual decomposition needs to be performed again for each distinct shifts.

Major computational savings may be obtained if C is low rank, namely $C = C_0R$, with $C_0 \in \mathbb{R}^{n\times\bar{m}}$ and $\bar{m} < m$. Indeed, the m shifted systems can be solved more efficiently by only working with the common matrix C_0. For the rest of this section we assume that C is full rank, and postpone the treatment of the low-rank case to later, when we discuss the occurrence of large B. Indeed, the rank of C is key in developing general projection methods, as explained next.

Projection methods. Let $\mathcal{V}$ be a space[6] of dimension k, and let the columns of $V_k \in \mathbb{C}^{n\times k}$ span $\mathcal{V}$. An approximate solution $\mathbf{X}_k$ with range$(\mathbf{X}_k) \subset \mathcal{V}$ is sought such that

$$R_k := A\mathbf{X}_k + \mathbf{X}_kB - C \approx 0.$$

Several options arise, depending on the choice of $\mathcal{V}$ and on the strategy to determine $\mathbf{X}_k$ within the space $\mathcal{V}$. For a given $\mathcal{V}$, let thus $\mathbf{X}_k = V_k\mathbf{Y}_k \approx \mathbf{X}$, for some $\mathbf{Y}_k \in \mathbb{R}^{k\times m}$ to be determined. Recalling the operator $\mathcal{S}$ defined in (4.7), we observe that $\mathcal{S}$ generalizes to the "block" B the concept of shifted matrices, namely

$$x \mapsto (A + \beta I)x = Ax + x\beta.$$

Therefore, it is very natural to extend the algorithmic strategies of linear systems to the case of $\mathcal{S}$. Extensions of the linear system solvers CG (FOM) and MINRES (GMRES) can be thought of for A symmetric (nonsymmetric), although the actual implementation differs. All these system solvers are derived by imposing some orthogonality condition on the system residual. Consider the matrix inner product defined as

$$\langle Y, X\rangle = \text{trace}(Y^*X). \tag{4.14}$$

[6] We use complex arithmetic for $\mathcal{V}$ to allow for complex spaces, which may arise when using Rational Krylov subspaces with complex shifts. A careful implementation can construct a real space in case conjugate shifts are used. For the sake of generality we stick to complex arithmetic for $\mathcal{V}$.

Algorithm. Given A, B, C
Orthogonalize the columns of C to get $v_1 = V_1$
for $k = 1, 2, \ldots$ **do**
  Compute $\mathbf{Y}_k$ solution to $(V_k^* A V_k)\mathbf{Y} + \mathbf{Y}B - V_k^* C = 0$
  if converged **then**
    $\mathbf{X}_k = V_k \mathbf{Y}_k$ and stop
  end if
  Arnoldi procedure for the next basis block:
    $\hat{v} = A v_k$
    Make $\hat{v}$ orthogonal wrto $\{v_1, \ldots, v_k\}$
    Orthogonalize the columns of $\hat{v}$ to get v_{k+1}
    Update: $V_{k+1} = [V_k, v_{k+1}]$
end for

FIG. 4.1. *Solution of Sylvester equation in $K_k^\square(A, C)$ with Galerkin condition*

If we require that the columns of R_k be orthogonal to the approximation space $\mathcal{V}$ in this inner product, then we are imposing a Galerkin condition, explicitly given by $V_k^* R_k = 0$. For simplicity let us assume that $V_k^* V_k = I$. Then

$$0 = V_k^* R_k = V_k^* A V_k \mathbf{Y}_k + \mathbf{Y}_k B - V_k^* C. \tag{4.15}$$

The condition thus gives a new, reduced in size, Sylvester equation to be solved. Under the hypothesis that $\mathrm{spec}(V_k^* A V_k) \cap \mathrm{spec}(-B) = \emptyset$, such equation can be solved efficiently by one of the methods discussed in section 4.2. The procedure above holds for any space $\mathcal{V}$ and associated full rank matrix V_k. Therefore, the effectiveness of the approximation process depends on the actual selection of $\mathcal{V}$. A well exercised choice is given by the block Krylov subspace

$$K_k^\square(A, C) = \mathrm{range}([C, AC, \ldots, A^{k-1}C]).$$

This space coincides with $K_k^\square(\mathcal{S}, C) = \mathrm{range}([C, \mathcal{S}(C), \ldots, \mathcal{S}^{k-1}(C)])$, where $\mathcal{S}^j(C) = \mathcal{S}(\mathcal{S}^{j-1}(C))$ and $\mathcal{S}^0(C) = C$ [215]. With this selection of space, the procedure just outlined is the complete analog of the one giving rise to the Full Orthogonalization Method (FOM) for $m = 1$ or for $B = 0$. However, due to possible loss of rank in the basis, it was suggested in [215] to generate the subspace with A rather than with $\mathcal{S}$. As an example, Figure 4.1 describes the algorithm implementing the projection method with the generation of the block Krylov subspace and the determination of the approximation by imposing the Galerkin orthogonality condition.

For later reference, we remark that the Arnoldi procedure used in Figure 4.1 generates a matrix recurrence that can be written as

$$A V_k = V_k H_k + \hat{v} e_k^\top, \tag{4.16}$$

where $\hat{v}$ is the new block of basis vectors, prior orthogonalization, and H_k contains the orthogonality coefficients, with $H_k = V_k^\top A V_k$.

One could consider constraint spaces different from the approximation spaces; in this case, a so-called Petrov-Galerkin condition is imposed on the residual. Following the linear system case, and using, e.g., the space spanned by the columns of AV_k, one would be tempted to impose the condition $(AV_k)^* R_k = 0$, giving

$$V_k^* A^* A V_k \mathbf{Y}_k + V_k^* A^* V_k \mathbf{Y}_k B - V_k^* A^* C = 0. \tag{4.17}$$

In the standard ($B = 0$) linear system setting, this condition is equivalent to minimizing the residual R_k in the Frobenius norm, that is

$$\min_{\mathbf{Y}_k \in \mathbb{R}^{k\times m}} \|R_k\|_F. \tag{4.18}$$

However, for $B \neq 0$, such equivalence does *not* hold, that is, the solution to (4.17) is not a residual minimizing approximation. To attain a residual minimization, the orthogonality condition should be applied to the operator $\mathcal{S}$ in (4.7); to this end, we note that the adjoint operator $\mathcal{S}^*$ with respect to the inner product in (4.14) is given by $\mathcal{S}^*(X) = A^*X + XB^*$.

PROPOSITION 4.2. [215, sec.3] *Let $\mathbf{Y}_k \in \mathbb{R}^{k\times m}$, and let $R_k = AV_k\mathbf{Y}_k + V_k\mathbf{Y}_kB - C$ be the associated residual. Then*

$$\mathbf{Y}_k = \arg\min_{\mathbf{Y}_k \in \mathbb{R}^{k\times m}} \|R_k\|_F \quad \text{if and only if} \quad V_k^*\mathcal{S}^*(R_k) = 0.$$

For the choice $\mathcal{V} = K_k^{\square}(A, C)$, the minimization process in (4.18) is the matrix analog of GMRES (for $m = 1$ or $B = 0$) (cf. [220, sec. 6.12]). Similar results are discussed independently in [111]. In [215] the authors provide a detailed description of the parallel between solving (4.11) for $m \ll n$ with Galerkin and with minimizing procedures, and solving linear systems $A\mathbf{X} = C$ by means of block methods. This confirms the intuitive fact that thanks to the small rank of B and to the fact that C is full rank, there is a sort of "block" shift invariance property taking place, where the "shift" block is the matrix B. Numerical experiments on the performance of the approach are also given in [215]. Upper bounds for the residual norm of Galerkin and residual minimizing methods with $\mathcal{V} = K_k^{\square}(A, C)$ are also provided in [215].

An iteration based on block Arnoldi was recently proposed in [149], where a series of Sylvester equations is solved by means of the block method above; however, the full rank of the right-hand side cannot be ensured after the first iteration of the process.

An alternative choice of approximation space $\mathcal{V}$ has recently shown great potential compared to the block Krylov subspace, and it is given by the *Extended* Krylov subspace, defined as $\mathbf{EK}_k(A, C) := K_k^{\square}(A, C) + K_k^{\square}(A^{-1}, A^{-1}C)$. Since the spaces are nested, namely $\mathbf{EK}_k(A, C) \subseteq \mathbf{EK}_{k+1}(A, C)$, the space can be generated iteratively, allowing one to improve the approximate solution as the recurrence proceeds. Experiments in [233] show that the good performance of the derived method seems to fully compensate the high costs of solving linear systems with a large and sparse A at each iteration.

4.4. Sylvester equation. Large A and large B. In the most general case, both A and B have large dimensions. This setting arises in many situations, as in the discretization of separable PDEs [83], or in the computation of the *cross* Gramian in control [4]. The possibility of writing $C = C_1C_2^\top$ with C_1, C_2 with low column rank is crucial to obtain good low-rank approximations to $\mathbf{X}$, thus avoiding the storage of the whole matrix, which is in general prohibitive.

Methods in the literature have mainly proceeded in three directions: Projection type approaches (mostly based on the Krylov subspace family), matrix updating sequences (such as Alternating-Direction-Implicit iterations), and sparse data format recurrences. Combinations of these have also been explored.

It is interesting that, due to the important role the Lyapunov equation has in control problems, many authors have developed numerical procedures specifically for the Lyapunov equation, and not for the Sylvester equation, although in many cases

they could be extended to the latter in a natural manner. For historical reasons, and also to avoid constant reference to the equation context, we will refer to the literature the way the methods were originally presented. In particular, it will be apparent that the literature on large scale Lyapunov equations is richer than that for the Sylvester equation.

We also notice that, as a major distinction from linear vector equations, the numerical solution of matrix equations cannot directly rely on preconditioning strategies, unless the Kronecker formulation is employed. Indeed, preconditioning methods would necessarily destroy the symmetry properties of the problem, and its peculiar Kronecker structure, which allows one to deal with computational costs that depend on powers of n and m, but not on powers of $n \cdot m$.

4.4.1. Projection methods. When both n and m are large, the dense solution matrix $\mathbf{X}$ of (4.1) cannot be stored, therefore the determination of a memory saving approximation becomes mandatory. Whenever $C = C_1 C_2^\top$ has low rank, the results discussed in section 4.1 suggest that a low rank approximate solution may be determined, so that projection strategies are very appealing. Indeed, these methods determine low rank approximations $\widetilde{\mathbf{X}} = V_k \mathbf{Y} W_j^\top \approx \mathbf{X}$, with V_k and W_j having far fewer columns than n and m, respectively, and are generalizations of the procedure seen in previous sections.

Let $\mathcal{V}$ and $\mathcal{W}$ be two subspaces, in principle not necessarily of the same dimension, and let the k (j) columns of V_k (of W_j) be orthonormal bases for $\mathcal{V}$ (for $\mathcal{W}$), with $k \ll n$, $j \ll m$, such that $\mathcal{V}$ is not orthogonal to range(C_1) and $\mathcal{W}$ is not orthogonal to range(C_2). We look for an approximation $\widetilde{\mathbf{X}} = V_k \mathbf{Y} W_j^\top \approx \mathbf{X}$, and we let $R := C_1 C_2^\top - A\widetilde{\mathbf{X}} - \widetilde{\mathbf{X}} B$ be the associated residual. Then we have $\widetilde{\mathbf{x}} = \text{vec}(\widetilde{\mathbf{X}}) = (W_j \otimes V_k)\text{vec}(\mathbf{Y})$, where $\widetilde{\mathbf{x}}$ is an approximate solution of (4.3). Imposing a Galerkin (orthogonality) condition to the vector residual $c - \mathcal{A}\widetilde{\mathbf{x}}$ with respect to the space spanned by $W_j \otimes V_k$ corresponds to writing

$$(W_j \otimes V_k)^\top (c - \mathcal{A}\widetilde{\mathbf{x}}) = 0 \quad \Leftrightarrow \quad V_k^\top R W_j = 0.$$

Alternatively, such Galerkin condition may be imposed by using matrix orthogonality with respect to the inner product in (4.14). Other conditions could be considered, such as the minimization of the residual in some norm, or the orthogonality of the residual with respect to some other space; see, e.g., [133], [129], [181]. Substituting the residual matrix, the equation $V_k^\top R W_j = 0$ gives the following small size Sylvester equation:

$$V_k^\top A V_k \mathbf{Y} + \mathbf{Y} W_j^\top B W_j = V_k^\top C_1 (W_j^\top C_2)^\top. \tag{4.19}$$

If $V_k^\top A V_k$ and $W_j^\top B W_j$ have disjoint spectra, then this equation admits a unique solution for any right-hand side. By assuming that the fields of values of A and B are disjoint, one can ensure that $V_k^\top A V_k$ and $W_j^\top B W_j$ have disjoint spectra. Though restrictive, such assumption is welcome also for stability purposes, so as to monitor that the solution $\mathbf{X}$ have moderate norm [262].

The quality of the approximation depends on the choice of $\mathcal{V}$ and $\mathcal{W}$, which is usually based on similar arguments for the two spaces. In his seminal article [219], Saad proposed Krylov subspaces for determining a low rank approximate solution to the Lyapunov equation by projection; the motivation was that Krylov subspaces tend to approximate well the action of the matrix exponential to a vector, so that the solution in the integral form (4.5) can take advantage of this property (see also section 5

for an explicit derivation). A major problem with this approach is that both bases V_k and W_j need to be stored to compute the final approximate solution. Since both matrices are full, this provides a severe limitation in the maximum affordable size of the two approximation spaces when A and B are large. In the quest of small but more effective spaces, several alternatives have been investigated. The impressive performance results of these enriched spaces have lead to a resurgence of projection-type methods for linear matrix equations. In addition to the standard Krylov subspace, we list here a couple of recently explored selections for $\mathcal{V}$ with A and C_1; similar choices can be made for $\mathcal{W}$ using $B^\top$ and C_2.

(a) Standard (block) Krylov subspace:

$$\mathcal{V} = \text{range}([C_1, AC_1, A^2C_1, \ldots]);$$

(b) Rational (block) Krylov subspace:

$$\mathcal{V} = \text{range}([(A-\sigma_1 I)^{-1}C_1, (A-\sigma_2 I)^{-1}(A-\sigma_1 I)^{-1}C_1, \ldots]),$$

for a specifically chosen sequence $\{\sigma_j\}$, $j = 1, 2, \ldots$ that ensure nonsingularity of the shifted matrix.

(c) Global Krylov subspace:

$$\mathcal{V} = \left\{ \sum_{i \geq 0} A^i C_1 \gamma_i,\ \gamma_i \in \mathbb{R} \right\} = \text{span}\{C_1, AC_1, A^2C_1, \ldots\},$$

where the linear combination is performed block-wise.

In all instances the least number of powers is computed so as to reach the dimension k. All spaces listed above are nested, so that an approximate solution can be derived while each of these spaces are expanded. The idea of generating different approximation spaces - of the same dimension - for A and B by means of standard Krylov subspaces was first developed in [133], where however the right-hand side C of the original problem was approximated by a rank-one matrix $c_1 c_2^\top$, to be able to build the standard Krylov subspaces $K_j(A, c_1)$ and $K_j(B^\top, c_2)$ as approximation spaces. The approach was then generalized to block Krylov subspaces in [231], so as to exploit the low (but possibly larger than one) rank matrices C_1, C_2. Different Krylov subspaces for the right and left subspace should be considered also when $B = A^\top$, as long as $C_1 C_2^\top$ is nonsymmetric. Nonetheless, in this case the generation of the two spaces can share some computationally intensive computation, such as shifted system solves with the same coefficient matrix. The possibility of using nonsymmetric Lanczos, processes which, e.g., simultaneously generate $K_j(A, C_1)$ and $K_j(A, C_2)$ could also be considered.

The subspaces listed above are somewhat related. For instance, the standard Krylov subspace can be formally obtained from the rational Krylov subspace for $\sigma_j = \infty$ for all j. Moreover, the Rational block Krylov subspace also includes the special choice of fixed poles at zero and infinity, which corresponds to the *extended* Krylov subspace $K_j^\square(A, C_1) + K_k^\square(A^{-1}, A^{-1}C_1)$, where j and k can in principle be different [76]. In addition, one can impose that C_1 belongs to the rational Krylov subspace with the choice $\sigma_1 = \infty$. The global Krylov subspace in (c) is a subspace of the block Krylov subspace; it was first proposed to solve linear systems with multiple right-hand sides [144], and then adapted to the Sylvester equation in [142]. Global spaces may be viewed as simplified versions of block Krylov spaces, where the polynomial coefficients

are chosen to be multiples of the identity matrices, therefore lowering the number of degrees of freedom.

Rational Krylov subspace have a rich history. First introduced by Ruhe in the context of eigenvalue approximation [218], its relevance has significantly spread in applied approximation theory and model order reduction frameworks, due to its functional approximation properties; see, e.g., [4],[109],[115] and references therein.

We should mention that the effectiveness of general rational spaces strongly relies on the efficiency of solving systems with A or its shifted variants. The reliability of recent direct sparse and iterative linear system solvers has made it possible to use these richer approximation spaces for more complex problems such as the ones we are addressing.

Already in [133] the possibility of restarting the process was considered. In this case, a maximum subspace dimension is allowed and the final approximate solution is obtained as $\widetilde{\mathbf{X}} = \widetilde{\mathbf{X}}^{(1)} + \widetilde{\mathbf{X}}^{(2)} + \widetilde{\mathbf{X}}^{(3)} + \ldots$, where the superscripts indicate a new restart. Strategies on how to generate the new approximations were proposed in [133]. We mention that new restarting procedures were recently proposed in [3], but their overall computational costs for large scale matrices have not clearly been assessed. An alternative that could be considered in the symmetric case is to resort to a two-pass strategy, inspired by a similar procedure in the eigenvalue context. Indeed, for A and B symmetric, not necessarily equal, an orthogonal basis of each standard Krylov subspace together with the projected matrix could be generated without storing the whole basis, but only the last three (block) vectors, because the orthogonalization process reduces to the short-term Lanczos recurrence [220]. Therefore, in a first-pass only the projected solution $\mathbf{Y}$ could be determined while limiting the storage for V_k and W_j; at convergence the approximate solution $\widetilde{\mathbf{X}} = V_k \mathbf{Y} W_j^\top$ could be recovered by generating the two bases once again, and updating $\widetilde{\mathbf{X}}$ on the fly with the already computed $\mathbf{Y}$; an implementation of such approach can be found in [162] for $B = A^\top$ and $C_1 = C_2$. The same idea could be used for other situations where a short-term recurrence is viable; the effectiveness of the overall method strongly depends on the affordability of computing the two bases twice.

The convergence analysis of projection methods has long been overlooked, in spite of a good experimental performance of these strategies. Following recent significant advances in the convergence of projection methods for the Lyapunov equation (see section 5.2.1) Beckermann in [21] provided a thorough study: residual norm bounds are given for Galerkin projection methods when Rational Krylov subspaces, of possibly different dimensions, are used for A and $B^\top$. The proposed estimates rely on new residual relations, and highlight the role of the field of values of the two coefficient matrices. Advances in the theoretical aspects of projection methods have been made in close connection with the recent great steps ahead taken in the understanding of polynomial and rational approximation methods for matrix functions such as the matrix exponential. The interplay of numerical linear algebra, approximation theory and functional analysis has made this possible; see, e.g., [20],[115],[125],[77],[156],[137],[116],[108],[82],[126],[194] and their references.

4.4.2. ADI iteration. The Alternating-Direction-Implicit (ADI) iteration was first introduced in [199] in 1955, and proposed to solve large Sylvester equations by Ellner and Wachspress in [83][7]. Since then, and with various computationally effective refinements, the approach has been one of the leading methods for solving

[7] The authors of [83] referred to these Sylvester equations as Lyapunov equations.

large-scale Sylvester (and Lyapunov) equations. In its original form discussed in [83] and summarized next, the ADI iteration is derived for a full matrix $\mathbf{X}$ (cf. also Smith [239] for the derivation below). A low memory factorized version is used in practice for large matrices, and it will be presented in the sequel. In the following we assume that both real matrices A and B have eigenvalues with positive real parts. We can equivalently rewrite (4.1) as

$$(qI + A)\mathbf{X}(qI + B) - (qI - A)\mathbf{X}(qI - B) = 2qC, \quad q \neq 0.$$

For $q > 0$, $qI + A$ and $qI + B$ are nonsingular and we can multiply by their inverses so as to obtain the Stein equation

$$\mathbf{X} - (qI + A)^{-1}(qI - A)\mathbf{X}(qI - B)(qI + B)^{-1} = 2q(qI + A)^{-1}C(qI + B)^{-1}.$$

Let $\mathcal{A} = (qI+A)^{-1}(qI-A)$, $\mathcal{B} = (qI-B)(qI+B)^{-1}$ and $\mathcal{C} = 2q(qI+A)^{-1}C(qI+B)^{-1}$. Then the matrix $\mathbf{X} = \sum_{k=1}^{\infty} \mathcal{A}^{k-1}\mathcal{C}\mathcal{B}^{k-1}$ is a formal solution, and since both $\mathcal{A}$ and $\mathcal{B}$ have spectral radius less than one, the series is convergent. This consideration drives the implementation of the following sequence of approximations

$$\mathbf{X}_0 = \mathcal{C}, \quad \mathbf{X}_{k+1} = \mathcal{C} + \mathcal{A}\mathbf{X}_k\mathcal{B} \tag{4.20}$$

The approach can be generalized to two parameters $p, q > 0$ for A and B, respectively, giving the transformed equation

$$\mathbf{X} - \mathcal{A}(p, q)\mathbf{X}\mathcal{B}(p, q) = \mathcal{C}(p, q),$$

with $\mathcal{A}(p, q) = (pI + A)^{-1}(A - qI)$, $\mathcal{B}(p, q) = (B - pI)(qI + B)^{-1}$ and $\mathcal{C}(p, q) = (p + q)(pI + A)^{-1}C(qI + B)^{-1}$. A recursion similar to the one for a single parameter can be derived, and it is convergent if the spectral radii of $\mathcal{A}(p, q), \mathcal{B}(p, q)$ are both less than one. Therefore, the parameters p, q are selected so as to minimize these spectral radii, and if A, B are both symmetric with spectral intervals (a, b) and (c, d), respectively, this corresponds to solving the ADI minimax problem

$$\min_{p,q>0} \max_{s\in(a,b),t\in(c,d)} \left| \frac{(q - s)(p - t)}{(p + s)(q + t)} \right|.$$

The generalization of this concept allows one to choose different p, q at each iteration, providing a sequence of parameters $p_1, p_2, \ldots$ and $q_1, q_2, \ldots$, and then cyclically repeating them until convergence. Following a successful idea developed for the Lyapunov equation, the authors of [27] propose a factorized version of the ADI iteration, which allows one to write the approximate solution as the product of three memory saving factors, as long as $C = C_1C_2^*$ is low rank. We will expand on this implementation aspect in the case of the Lyapunov equation. In that setting, we shall also emphasize the tight connection between ADI and projection methods.

4.4.3. Data sparse and other methods. The Kronecker formulation allows one to consider a wide range of linear system solvers for (4.3); an early ad-hoc implementation of the classical SOR was proposed in [244], although the exploding dimensions of the problem significantly penalize the method, when compared with the approaches analyzed so far. We also recall from section 4.4.1 that global Krylov subspace methods represent an implicit way to deal with the Kronecker formulation. Other iterative solvers based on the Kronecker formulation (4.3) have been explored

specifically for the Lyapunov equation, and they will be reviewed in section 5.2.3. These appear to be the main directions taken whenever C is not numerically low rank.

For data-sparse matrices $A \in \mathbb{R}^{n\times n}$ and $B \in \mathbb{R}^{m\times m}$, namely such that matrix-vector multiplications for A and B can be performed with complexity $\mathcal{O}(n)$ and $\mathcal{O}(m)$, respectively, an extension of the multigrid algorithm was recently proposed in [104], where A and B stem from the discretization of a class of partial differential equations, and their spectra are assumed to be separated by a line. A particular computational caveat of this extension is the smoother (e.g., Jacobi), which in this case requires approximately solving a *diagonal* Sylvester equation at each iteration. This step is carefully discussed in [104], and a procedure for determining a cheap and low rank approximate solution is devised. Other crucial points include handling the connection between the (independently generated) sequences of matrices for A and B, which is accounted for during the smoothing procedure, and the imposition of regularity constraints on the continuous operators associated with A and B. A major issue arising when using these hierarchical methods is whether the approximate solution $\widetilde{\mathbf{X}}$ is low rank, so that it can be stored cheaply by means of a (hierarchical) sparse format, the $\mathcal{H}$-matrix format. Such format is a data-sparse representation for a special class of matrices, which appear to occur after the use of several discretization methods, when PDEs or integral equations are treated numerically [106]. The $\mathcal{H}$-matrix format consists of partitioning a given matrix recursively, into submatrices admitting low-rank approximations. The definition of this format requires the introduction of further arithmetic operations/approximations, so as to be able to determine, e.g., an $\mathcal{H}$-matrix after the approximate inversion of an $\mathcal{H}$-matrix, so as to make the class closed with respect to some important matrix operations; we refer to section 5.2.3 for further details.

A different though related approach consists in adapting small scale iterations to the large setting, again under the condition that C is low rank. This can be performed, for instance, within the sign function iteration, by using rank truncation of the iterates, and sparse format for the approximate solution. More details on the sign function iteration will be given in section 5.2.3. Here we mention that such an approach is investigated in [17] (see also [19]), where the sparse format chosen for the data and for the approximate solution is the hierarchical $\mathcal{H}$-matrix format also used in [102], [104]. With this approach, sparse approximate solutions to a Sylvester equation of size up to $n = 262144$ associated with a control problem for the two-dimensional heat equation are reported in [17]. The accuracy and effectiveness of the method depend on some thresholds used for maintaining sparsity and low rank during the iteration, and are thus problem dependent.

5. Continuous-time Lyapunov equation. For $B = A^\top$ we obtain a simplified version of the Sylvester equation, the Lyapunov equation

$$A\mathbf{X} + \mathbf{X}A^\top = C, \tag{5.1}$$

with C symmetric, and its generalized counterpart,

$$A\mathbf{X}E^\top + E\mathbf{X}A^\top = C \tag{5.2}$$

with E nonsingular. Clearly, this latter equation can be transformed into the form (5.1) by left and right multiplication by E^{-1} and $E^{-\top}$, respectively. If E is symmetric and positive definite, a Cholesky decomposition could be performed and its

inverse factors applied to the equation on the left and right sides, so as to maintain the problem structure. These are called the *continuous-time* Lyapunov equations, to be distinguished from the discrete-time equations, which will be discussed in section 6. They arise in the analysis of continuous-time and discrete-time linear dynamical systems, respectively. A very detailed analysis of the Lyapunov equation, with computational developments until 1995 and many relevant connections in the control application area can be found in [91].

In the context of inertia theory, (5.1) with $C \geq 0$ relates the location of the eigenvalues of both A and $\mathbf{X}$ with respect to the imaginary axis. Since C is symmetric, the solution $\mathbf{X}$ is also symmetric. According to the Sylvester equation theory, the solution to (5.1) exists and is unique if and only if $\lambda_i + \bar{\lambda}_j \neq 0$ for all eigenvalues λ_i, λ_j of A [132]. If all eigenvalues of A have negative real part, namely A is stable, then such condition is satisfied, so that a unique solution is ensured. We remark that the stability of A is an important property in the control setting, therefore it is not regarded as a restriction for solving the Lyapunov equation, although not strictly required. We shall see, however, that some of the large-scale methods require additional restrictions on A, namely its passivity, to ensure convergence. For A nonsymmetric, this extra condition may limit the applicability of the method, since in general a stable matrix A is not necessarily passive. It can be verified that if A is stable and $C > 0$ ($C \geq 0$) then $\mathbf{X} > 0$ ($\mathbf{X} \geq 0$); in this case the problem is called the *stable* Lyapunov equation. If $C \geq 0$ and $(A, C^\top)$ is observable, then $\mathbf{X} > 0$. A detailed account on various relations between the inertia of A and of $\mathbf{X}$ can be found, e.g., in [169, section 13.1], [227],[228]. A specialized sensitivity bound can be obtained for the stable Lyapunov equation. Assume that $\mathbf{X} + \Delta\mathbf{X}$ solves

$$(A + \Delta A)(\mathbf{X} + \Delta\mathbf{X}) + (\mathbf{X} + \Delta\mathbf{X})(A + \Delta A)^\top + (C + \Delta C) = 0$$

then

$$\frac{\|\Delta\mathbf{X}\|}{\|\mathbf{X} + \Delta\mathbf{X}\|} \leq 2\|A + \Delta A\|\,\|H\| \left[\frac{\|\Delta A\|}{\|A + \Delta A\|} + \frac{\|\Delta C\|}{\|C + \Delta C\|}\right],$$

where H satisfies $AH + HA^\top + I = 0$, and all denominators are assumed to be nonzero [122]. Estimates for the backward error associated with the Lyapunov equation do not differ from those in (4.8) for the Sylvester equation; therefore, except for the substitution $B = A^\top$, the extra structure of the problem does not modify the sensitivity properties of the solution [124].

The sensitivity of the solution to (5.1) can also be analyzed by looking at the spectral properties of the solution matrix; this topic has attracted a lot of interest, especially in light of its consequences in the stability analysis of dynamical systems. Various authors have explored the spectral decomposition of the Lyapunov solution, to make statements on the decay of its eigenvalues; see, e.g., [204],[240],[5],[158]. In [5], an error estimate for a low rank approximation to the solution of (5.1) was proved. For the sake of simplicity we report here only the case when C is rank-one. The result relies on the fact that the solution matrix admits the following decomposition:

$$\mathbf{X} = ZDZ^\top, \quad D = \operatorname{diag}(\delta_1, \ldots, \delta_n), \quad \delta_k = \frac{-1}{2\Re(\lambda_k)} \prod_{j=1}^{k-1} \left|\frac{\lambda_k - \lambda_j}{\lambda_k^* + \lambda_j}\right|^2,$$

where λ_j are the eigenvalues of the diagonalizable matrix A.

THEOREM 5.1. *Assume A is diagonalizable with eigenvector matrix Q, having all unit norm columns, and let $C = cc^\top$. Let $\mathbf{X} = \sum_{j=1}^n \delta_j z_j z_j^\top$ solve (5.1), with*

the nonnegative values δ_j sorted decreasingly, and for $k \in \{1, \dots, n\}$ define $\mathbf{X}_k = \sum_{j=1}^{k} \delta_j z_j z_j^\top$. *Then*

$$\|\mathbf{X} - \mathbf{X}_k\| \le (n-k)^2 \delta_{k+1} (\kappa_2(Q)\|c\|_2)^2,$$

where $\|\ \|$ is the matrix norm induced by the vector 2-norm.

The bound may not be sharp for highly non-normal A, for which $\kappa_2(Q)$ may be large. A more specialized bound was given by Penzl for A symmetric, which only depends on the condition number of A [204]. In fact, it was shown in [204] that eigenvalues alone cannot predict the eigenvalue distribution of $\mathbf{X}$; this is reminiscent of similar limitations of non-normal matrices in the convergence analysis of iterative methods for linear systems [179, sec. 5.7.3].

Bounds on the eigenvalue decay that attempt to cope with non-normality were obtained in [221, sec.3.1.2], where the concept of pseudospectrum is used; in there, some interesting counter-intuitive convergence behaviors are also described. Overall, the analysis of the decay in the solution spectrum is a largely open topic of research in the non-normal case.

In addition to the application relevance, establishing conditions under which the solution matrix has exponentially decaying eigenvalues provides theoretical motivation for the good performance of projection methods in the large scale case; indeed, these strategies aim at determining low rank approximations by a judicious choice of the approximation space.

5.1. Lyapunov equation. Small scale computation. As for the Sylvester equation, the closed form solutions described in section 4 could be used in theory. A detailed account of early methods can be found in [91], together with some ad-hoc algorithms appropriate when special forms of A (e.g., Schwarz, Companion or Jordan forms) are available; see also [40] for an improved approach for the companion form.

The standard method for efficiently solving (5.1) when A has small dimensions, does not essentially differ from those for the Sylvester equation discussed in previous sections. In fact, due to the fact that $B = A^\top$, the computational cost of the reduction to Schur form is halved in the Bartels-Stewart method [229].

A specifically designed algorithm was proposed by Hammarling, to exploit the case when C is positive semidefinite. It was shown in [117] that if $C = C_1 C_1^\top \ge 0$, it is possible to determine the Cholesky factor of the solution $\mathbf{X}$, that is L in $\mathbf{X} = LL^*$, without first determining $\mathbf{X}$. The computation of the Cholesky factor has some advantages when $\mathbf{X}$ is nonsingular but severely ill-conditioned, so that dealing with L significantly improves the accuracy and robustness of computations with $\mathbf{X}$; we refer to [293] for a comparison between Hammarling's and Bartel-Stewart methods. A block variant of the Hammarling's method for the discrete-time Lyapunov equation is suggested in [161], which can dramatically improve the performance of the original scalar (unpartitioned) algorithm on specific machine architectures, while preserving the stability of the original method.

We also mention the possibility of a pre-processing, proposed both in the continuous and discrete-time equations, so as to transform the original symmetric problem onto a skew-symmetric one, so that the solution will also be skew-symmetric ($\mathbf{X} = -\mathbf{X}^\top$), allowing for some memory savings; see [91, sec. 2.1.2] and references therein.

A completely different approach exploits the fact that the solution to the Lyapunov equation may be computed by means of matrix functions, in particular, by

using the sign function. Although less general than Schur-form-based algorithms, they allow one to handle larger problems, especially if the right-hand side is low-rank or structured, and can be more easily adapted to a high performance computational environment. The idea is to use well-established matrix iterations to obtain the matrix sign function in a cheap manner, by fully exploiting the possible sparse format of the matrix. The whole procedure is actually more general, and it applies to the quadratic algebraic Riccati equation, of which the Lyapunov equation is the linear portion, obtained by zeroing the second order term. Here we will follow the derivation proposed in [18], see also [29], although the main iteration was introduced by Larin and Aliev in [171] for the generalized Lyapunov equation. Let $\mathcal{A} = \mathcal{X}\text{blkdiag}(J_+, J_-)\mathcal{X}^{-1}$ be the Jordan decomposition of a given matrix $\mathcal{A}$, where J_+, J_- represent the Jordan matrices associated with the eigenvalues in the open planes $\mathbb{C}^+$ and in $\mathbb{C}^-$, respectively. Then $\text{sign}(\mathcal{A}) = \mathcal{X}\text{blkdiag}(I, -I)\mathcal{X}^{-1}$, where the dimensions of I, $-I$ match those of J_+ and J_-, respectively. For A stable, the solution to the Lyapunov equation satisfies (cf., e.g., [216], [90])

$$\begin{bmatrix} 0 & \mathbf{X} \\ 0 & I \end{bmatrix} = \frac{1}{2}\left(I + \text{sign}\left(\begin{bmatrix} A^\top & C \\ 0 & -A \end{bmatrix}\right)\right) =: \frac{1}{2}\left(I + \text{sign}\left(Z_0\right)\right). \tag{5.3}$$

With this property, the following matrix iteration corresponds to applying the Newton method to the nonlinear equation $(\text{sign}Z_0)^2 = I$:

$$Z_{k+1} = \frac{1}{2}(Z_k + Z_k^{-1}), \qquad k = 0, 1, \ldots, \tag{5.4}$$

yielding

$$\text{sign}\, Z_0 = \lim_{k\to\infty} Z_k = \begin{bmatrix} -I & 2\mathbf{X} \\ 0 & I \end{bmatrix}.$$

Although the iteration is globally and (asymptotically) quadratically convergent, the basic iteration above may have slow initial convergence, therefore it is often accelerated by using a parameterized procedure, that is $Z_{k+1} = \frac{1}{2}(c_k Z_k + (c_k Z_k)^{-1})$, $k = 0, 1, \ldots$, for an appropriate selection of the parameter $c_k > 0$ (see, e.g., [52]). A popular choice is $c_k = |\det(Z_k)Z|^{-\frac{1}{n}}$ [52], [29]; see also [7] for a review of other choices.

5.2. Lyapunov equation. Large scale computation. For large scale problems, the solution to (5.1) is usually not determined exactly, but it is only approximated. A particularly important observation is that A and $\mathbf{X}$ have the same dimensions, and that although A may be sparse, $\mathbf{X}$ is dense, in general. Many procedures exist to numerically solve (5.1), and they all try to determine a memory saving and computationally appealing approximation. This is achieved in most cases by looking for a *low* rank approximation $\widetilde{\mathbf{X}}$ to the solution of the stable problem, that can be written as $\widetilde{\mathbf{X}} = \mathbf{Z}\mathbf{Z}^*$, so that only the tall matrix $\mathbf{Z}$ is actually computed and stored. This is possible if, for instance, the right-hand side has low rank, since in that case we also have $\mathbf{X} \geq 0$. Strategies to approximate the general right-hand side by low-rank matrices have also been explored in the literature; cf., e.g., [271] [133]. In fact, for C definite, low-rank solutions may be hard to find. For instance, the equation $A\mathbf{X} + \mathbf{X}A^\top = I$ with A symmetric admits the unique solution $\mathbf{X} = \frac{1}{2}A^{-1}$, which is obviously full rank, with not necessarily exponentially decreasing eigenvalues.

To help fully grasp the relevance of the topic, we notice that a number of recent PhD theses have been devoted to the theory and computational aspects of the large-scale Lyapunov matrix equation, whose results have significantly advanced knowledge

on the problem; among them, [202],[196],[190],[130],[293],[260],[176],[221]. The list would expand if one were to also include closely related theses on model order reduction of linear dynamical systems.

We conclude this section by noticing that a systematic numerical comparison of all iterative methods described in the following subsections on a variety of very large problems (of size $n \gg 10^4$) is still lacking, although in our presentation some guidelines are given about the settings in which each of the discussed method should be preferred.

5.2.1. Projection-type methods. As in the case of the Sylvester equation, the derivation of a projection method can be determined by imposing, e.g., the Galerkin condition to the residual, with respect to some approximation space. In particular, from equation (4.19) with $k = j$, $V_k = W_j$ and $C_2 = C_1$, we obtain the following projected small size Lyapunov equation:

$$V_k^* A V_k \mathbf{Y} + \mathbf{Y} V_k^* A^\top V_k = V_k^* C_1 (V_k^* C_1)^* \tag{5.5}$$

whose solution matrix $\mathbf{Y}_k$ gives $\mathbf{X}_k = V_k \mathbf{Y}_k V_k^* \approx \mathbf{X}$. To ensure that (5.5) admits a unique solution, the matrix $V_k^* A V_k$ is assumed to be stable. Such sufficient condition is met in exact precision arithmetic by requiring that A be passive, which is the usual hypothesis when using projection methods. Such condition represents a limitation of projection methods, since the original problem admits a unique solution even in case of a stable but not necessarily passive A. On the other hand, these are sufficient conditions: projection methods can work in practice without such assumption, although may break down or show some erratic convergence behavior; see [181] for an analysis.

The actual solution does not need to be explicitly stored. In fact, since $\mathbf{Y}_k$ is positive semi-definite and numerically singular, it is possible to perform a truncated decomposition of $\mathbf{Y}_k$ as $\mathbf{Y}_k = LL^*$, so that only the slim factor $Z_k = V_k L$ of the solution $\mathbf{X}_k = Z_k Z_k^*$ needs to be stored.

It is interesting to observe that an apparently different (functional) approach, based on the approximation to the matrix exponential and on (4.4), leads exactly to the same approximation procedure. Indeed, the action of the matrix exponential to a vector, $\exp(tA)C_1$, can be approximated in the space $\mathcal{V}$ as $V_k \exp(tH_k)(V_k^* C_1)$ where $H_k = V_k^* A V_k$, so that the analytic expression in (4.4) for the solution can be approximated explicitly; this is in fact the way the Galerkin approximate solution was originally obtained in [219] for a rank-one matrix C_1; see also [90].

PROPOSITION 5.2 ([219]). *Let $\mathcal{V}$ be a subspace of $\mathbb{R}^n$, and let V be such that $\mathcal{V} = \mathrm{range}(V)$. Let also H be the projection and restriction of A onto $\mathcal{V}$, and $y_m = \exp(tH)(V^* C_1)$. Then the matrix $V\mathbf{Y}V^*$ with*

$$\mathbf{Y} = \int_0^\infty y_m(t) y_m(t)^* dt$$

is the Galerkin approximate solution to the Lyapunov equation in $\mathcal{V}$.

The procedure above is very general, and the success of the approach, in terms of computational cost, depends on the choice of the approximation space $\mathcal{V}$. All choices discussed in section 4.3 have been explored. For instance, the block Krylov subspace $K_k^\square(A, C_1)$ was exploited in [139], and it was referred to as the Arnoldi method, after the procedure used to build the block Krylov subspace. In [139], the following relation was shown for the residual $R_k = A\mathbf{X}_k + \mathbf{X}_k A^* + C_1 C_1^*$ ([139, Th. 2.1])

$$\|R_k\|_F = \sqrt{2}\|(v_{k+1}^* A v_k) E_k^* \mathbf{Y}_k\|_F, \qquad E_k^* = [0_m, \ldots, 0_m, I_m],$$

where v_{k+1} contains the next block of basis vectors. This relation represents a major computational saving with respect to other procedures, since it allows one to avoid the explicit and expensive computation of R_k, to be able to evaluate its norm.

Finally, in [139] it was also shown that the solution $\mathbf{X}_k$ is the exact solution to the nearby problem

$$(A - \Delta)\mathbf{X} + \mathbf{X}(A - \Delta)^* - C_1 C_1^* = 0$$

with $\Delta = V_{k+1}(V_{k+1}^* A V_k)V_k^*$, $\|\Delta\|_F = \|V_{k+1}^* A V_k\|_F$.

The asymptotic convergence of the Arnoldi method was recently analyzed in [234]. Here we report a sample of such analysis, which applies to A symmetric and positive definite, and C_1 of rank one and unit norm; the derived bound was shown in [234] to provide an accurate worst-case convergence rate of the method.

THEOREM 5.3. *Let A be symmetric and positive definite, and let $\lambda_{\min}$ be the smallest eigenvalue of A. Let $\hat{\lambda}_{\min}$, $\hat{\lambda}_{\max}$ be the extreme eigenvalues of $A + \lambda_{\min} I$ and $\hat{\kappa} = \hat{\lambda}_{\max}/\hat{\lambda}_{\min}$. Let $\mathbf{X}_m$ be the Galerkin approximate solution to $\mathbf{X}$ in a subspace of dimension m. Then*

$$\|\mathbf{X} - \mathbf{X}_m\| \leq \frac{\sqrt{\hat{\kappa}} + 1}{\hat{\lambda}_{\min}\sqrt{\hat{\kappa}}} \left(\frac{\sqrt{\hat{\kappa}} - 1}{\sqrt{\hat{\kappa}} + 1} \right)^m, \tag{5.6}$$

where the matrix norm is the one induced by the vector 2-norm.

This result shows that the error norm is bounded by the same quantity as for the Conjugate Gradient method applied to a standard linear system with coefficient matrix $A + \lambda_{\min} I$.

As already mentioned, the algorithmic steps to compute an approximate solution by projection remain unchanged when a different approximation space is used. In [232] an efficient method based on the extended Krylov subspace $\mathbf{EK}_k(A, C_1) = K_k^{\square}(A, C_1) + K_k^{\square}(A^{-1}, A^{-1}C_1)$ was introduced, for C_1 of low rank. The proposed algorithm is very similar to the Arnoldi method, and a recurrence is proposed to derive the projection and restriction matrix $V_k^* A V_k$, where the columns of V_k span $\mathbf{EK}_k(A, C_1)$. In [232] the method was experimentally shown to be highly superior to the Arnoldi method for sparse and large coefficient matrices, allowing the computation of an equally accurate solution with a significantly smaller dimensional subspace, at lower computation costs. According to the experiments in [232], the method compares rather well also with respect to the Alternating-Direction-Implicit method. A recent asymptotic analysis in [157] theoretically confirmed these results, showing that the Extended Krylov subspace method is expected to have higher convergence rate than the Arnoldi method, and than ADI with a single pole.

More recently, a rational Krylov subspace method was employed as approximation space, which showed promising performance. As already mentioned, the general rational Krylov subspace requires the selection of a series of shifts (poles), which can either be computed a-priori or during the generation of the space. The a-priori pole computation may require possibly high computational costs, following the same procedure as that used for other parameter-dependent methods such as ADI (see below). On the other hand, it was shown in [78] that it is possible to employ a greedy algorithm to compute the next pole on the fly, while the iteration proceeds, with no extra $\mathcal{O}(n)$ computation. This is done by exploiting approximate spectral information generated within the current approximation space. Numerical experiments reported in [78] show that the method is superior to the Extended Krylov subspace when, for

instance, the field of values of A is very close to the imaginary axis. On the other hand, the computational cost of the general rational Krylov subspace method may be much higher, since a new shifted linear system needs to be solved at each iteration. However, the numerical experiments reported in [78] seem to ensure that the rational approximation space dimension remains usually very low, so that few systems have to be solved. The rational function idea is particularly appealing when $C = C_1 C_1^\top$ has rank p larger than one. In that case, the extended Krylov subspace increases its dimension by $2p$ vectors per iteration, making the whole procedure memory consuming if convergence is not fast. In [79] a tangential procedure is proposed to expand the block rational Krylov subspace at each iteration, so that only the most relevant directions are retained. More precisely, small matrices $d_1, \ldots, d_k$ are determined so that the following space is constructed

$$\text{range}([(A + s_1 I)^{-1} B d_1, (A + s_2 I)^{-1} B d_2, \ldots, (A + s_k I)^{-1} B d_k]).$$

Numerical experiments reported in [79] show that this strategy is capable of successfully handling the presence of many columns in C_1.

The Global Krylov subspace method for the Sylvester equation was applied to the Lyapunov equation in [145], with natural simplifications due to the fact that one single space needs to be generated; numerical experiments in [145] showed better performance than the standard block Krylov subspace methods. We also refer to [123] for a generalization to the simultaneous solution of a coupled pair of Lyapunov equations, corresponding to the two Gramians of a dynamical system. This last problem was also considered in [138]: the coupled block Arnoldi method and the nonsymmetric block Lanczos recurrence were analyzed for simultaneously approximating both Gramians, so as to obtain approximations to the linear transfer function of the system; see also [140] for enhancements of the proposed approaches.

As for the equations in previous sections, the Galerkin condition for the residual can be replaced by a Petrov-Galerkin condition. If the constraint space is $A\mathcal{V}$, then the resulting algorithm minimizes the residual in the Frobenius norm. This approach was explored in [139] for the standard block Krylov subspace and in [133] for the rank-one case. The projected problem entails the solution of a matrix linear least squares problem, for which expensive procedures were proposed [139],[133]. More recently, the minimal residual method was revisited and a more effective solver for the inner problem was proposed [181].

5.2.2. ADI method. For $B = A^\top$, the ADI method of section 4.4.2 for the Sylvester equation simplifies, leading to the following recursion with two half steps (cf. [269]):

$$\begin{aligned} \mathbf{X}_0 &= 0, \\ (A + s_j I)\mathbf{X}_{j-\frac{1}{2}} &= C_1 C_1^\top - \mathbf{X}_{j-1}(A^\top - s_j I_n), \\ (A + s_j I)\mathbf{X}_j &= C_1 C_1^\top - (\mathbf{X}_{j-\frac{1}{2}})^*(A^\top - s_j I_n), \quad j = 1, \ldots, k. \end{aligned}$$

Here the shifts $\{s_1, s_2, \ldots\}$ are complex, and are employed cyclically. If both A and C_1 are real, then the approximate solution will be real and symmetric, as long as both complex conjugates are used as shifts when complex shifts arise [184]. This implementation suffers from the fact that the whole matrix $\mathbf{X}_j$ needs to be stored, limiting its applicability to low dimensional problems. A key idea to make the recursion amenable to large dimension matrices is to keep the iterate in factored form.

This idea was successfully explored by Penzl in [203] and was the basis for the software package `Lyapack` [205]; see also [36]. The resulting low-rank ADI (LR-ADI) thus determines a recurrence for the factor $\mathbf{Z}_j$ of $\mathbf{X}_j = \mathbf{Z}_j\mathbf{Z}_j^*$ as

$$\mathbf{Z}_{j+1} = [(A^* - s_jI)(A^* + s_jI)^{-1}\mathbf{Z}_j,\ \sqrt{-2s_j}(A^* + s_jI)^{-1}C_1], \tag{5.7}$$

with $\mathbf{Z}_1 = \sqrt{-2s_1}(A^* + s_1I)^{-1}C_1$. Therefore, the number of columns in the factor $\mathbf{Z}_j$ is enlarged by $\mathrm{rank}(C_1)$ columns at each iteration. The success of the low-rank approach is related to what Penzl called the *low-rank phenomenon* in the solution $\mathbf{X}$, namely the already mentioned fact that the eigenvalues of $\mathbf{X}$ tend to decay very quickly towards machine precision, so that a low rank approximation appears to be possible (cf. section 4.1).

The iteration matrix $\mathbf{Z}_j$ is complex during the whole iteration, if some of the shifts are complex. A way to overcome this problem and to maintain real arithmetic throughout whenever A and C_1 are real, is discussed in [203]; see also the more recent contribution [26].

The iteration in (5.7) requires solving systems with right-hand sides $\mathbf{Z}_j$ and C_1 at each step j. A computational improvement to decrease the number of solves per iteration was suggested in [175] (where the LR-ADI method was called CF-ADI, in which CF stands for Cholesky Factor). In there, the columns were reordered so that only the previous iterate requires solves with a shifted matrix. The resulting recurrence is given by (cf. [175, Algorithm 2]):

$$\mathbf{z}_j = \frac{\sqrt{-2s_{j+1}}}{\sqrt{-2p_j}}(I - (s_{j+1} + s_j)(A + s_{j+1}I)^{-1})\mathbf{z}_{j-1}, \qquad \mathbf{Z}_j = [\mathbf{Z}_{j-1},\ \mathbf{z}_j],\ j = 2, 3, \ldots,$$

with $\mathbf{z}_1 = \sqrt{-2s_1}(A + s_1I)^{-1}C_1$ and $\mathbf{Z}_1 = \mathbf{z}_1$.

Additional recent contributions were devoted to the improvement of the computational costs per iteration. A strategy for reducing the number of solves was proposed under the name of "modified" low-rank Smith method in [113]. The idea is to compute the singular value decomposition of the iterate at each step and, given a dropping tolerance, to replace the iterate with its best low-rank approximation. A main ingredient is that the SVD is not recomputed from scratch; instead, it is updated after each step to include the new information and then truncated to retain only those singular values that lie above the specified tolerance. The use of the SVD exploits the fact that if $Z \approx V\Sigma U^*$ is a truncated SVD decomposition of Z, then $\mathbf{X} = ZZ^* \approx V\Sigma^2V^*$ is the truncated spectral decomposition of $\mathbf{X}$, so that the low rank factor can be readily maintained. In general, the procedure reduces the number of system solves per iteration, in a way that depends on the linear independence of the new iterate columns with respect to those of previous steps.

A bound for the difference between the traces of the solution $\mathbf{X}$ of Lyapunov equation and its ADI approximation is proposed in [255], which shows that the right-hand side of the Lyapunov equation can sometimes greatly influence the eigenvalue decay rate of the solution.

Computation of the shifts. The selection of the ADI parameters and their number have been a major topic of research for many years, since the performance of the method, in terms of number of iterations, heavily depends on them.

Let A be stable. Assuming a zero starting approximate solution, from the general ADI recurrence it follows that the error matrix associated with the ADI approximation

$\mathbf{X}_k^{\mathrm{ADI}}$ after k full iterations is given by (see also [203])

$$\mathbf{X} - \mathbf{X}_k^{\mathrm{ADI}} = (\bar{r}_k(A) r_k(-A)^{-1}) \mathbf{X} \bar{r}_k(A)^* r_k(-A)^{-*}, \tag{5.8}$$

with[8]

$$r_k(z) = \prod_{i=1}^{k} (s_i - z), \qquad \bar{r}_k(z) = \prod_{i=1}^{k} (\bar{s}_i - z).$$

This expression shows that for A diagonalizable, optimal parameters can be obtained by solving the minimax problem

$$\min_{s_1,\ldots,s_k \in \mathbb{C}^-} \max_{\lambda \in \Lambda(A)} \prod_{j=1}^{k} \left| \frac{\lambda - s_j}{\lambda + s_j} \right|. \tag{5.9}$$

The value of k is adjusted so that the set $\{s_1, \ldots, s_k\}$ is closed under conjugation in the case A is real. For A having real spectrum, this minimax problem was solved by Zolotaryov; if A is also symmetric, this leads to asymptotically optimal linear convergence rate for the approximation. The optimal parameters are given as (see, e.g., [268],[84])

$$s_j = \mathrm{dn}\left[\frac{(2j-1)K}{2t}, m \right], \quad j = 1, \ldots, k$$

where dn is a Jacobian elliptic function, and K is the complete elliptic integral of the first kind, of modulus m [1]. Generalizations to the case when the complex spectrum lies in certain specified complex regions Ω were discussed in [84]. However, it was only with the heuristic approach of Penzl in [203] that the computation of *suboptimal* ADI parameters became a more manageable procedure. The proposed strategy is performed as a pre-processing of the actual ADI computation: consider the Krylov subspaces $K_{k_A}(A, c)$, $K_{k_{A^{-1}}}(A^{-1}, c)$, and let V, W be such their orthonormal columns span the two spaces, respectively. Let Ω_+, Ω_- be the regions containing the eigenvalues of $V^\top A V$ and of $W^\top A W$ (the Ritz values). The key idea in [203] is to replace the spectrum of A with the region $\Omega := \Omega_+ \cup \Omega_-$, and then solve the minimax problem (5.9). The set Ω may be regarded as a reasonable approximation to the region of interest, the convex hull of the spectrum of A, and it can be more cheaply computed, especially for small $k_A, k_{A^{-1}}$; we refer to [203] and to the package [205] for more technical details.

If the data are real, all ADI computation can be carried out in real arithmetic, even if complex (conjugate) shifts are used [26]. In spite of the significant improvements in the ADI optimal parameter estimation, however, the method remains quite sensitive to the choice of these shifts, and performance can vary dramatically even for small changes in $k_A, k_{A^{-1}}$; see, e.g., the experiments in [232]. Adaptive strategies for pole selections such as those derived for the rational Krylov subspace in [78] are hard to obtain, since a basis for the generated space is not available. Nonetheless, these considerations have led to the investigation of hybrid approaches, which are described later in this section.

It was observed in [175] that the ADI method actually generates a (block) rational Krylov subspace for the given vector of shifts $\mathbf{s}_k = [s_1, \ldots, s_k]$. The connection

[8] Here and in the following $\bar{s}$ is the complex conjugate of s.

between the ADI method and the Galerkin rational Krylov subspace $\mathcal{K}_k(A, C_1, \mathbf{s}_k) = \text{range}((A - s_1 I)^{-1} C_1, \ldots, (A - s_k I)^{-1} C_1\}$, can be made more precise, when the two methods are used with the same parameters.

THEOREM 5.4. ([80, Th.3.4]) *Assume that the field of values of A and s_j, $j = 1, \ldots, k$ lie in the same half complex plane, and that C_1 has rank one. Let the columns of V form an orthonormal basis of $\mathcal{K}_k(A, C_1, \mathbf{s}_k)$, and let λ_j, $j = 1, \ldots, k$ be the Ritz values of A onto $\mathcal{K}_k(A, C_1, \mathbf{s}_k)$, that is λ_j are the eigenvalues of $V^* A V$. Then the ADI approximation coincides with the Galerkin approximate solution $\mathbf{X}_k$ with $K_k(A, C_1, \mathbf{s}_k)$ if and only if $s_j = \overline{\lambda_j}$, $j = 1, \ldots, k$ (under a suitable index permutation for the λ_j's).*

The condition $s_j = \overline{\lambda_j}$, $j = 1, \ldots, k$ is seldomly satisfied when the shifts are obtained by either an adaptive procedure or by a Penzl-style pre-processing (however, see [112] for an iterative process that approximates such a set of parameters, in the context of optimal model order reduction). Moreover, the approximation error in the Galerkin rational Krylov subspace approach may be bounded in terms of the same bound used for ADI. Since for the latter the upper bound can be reached, these results indicate that the projection onto a rational Krylov subspace provides better results than ADI, in terms of number iterations, if the same shifts are used.

Hybrid ADI methods. It was observed in [30] that "The most criticized property of the ADI iteration for solving Lyapunov equations is its demand for a set of good shift parameters to ensure fast convergence. [...] Most of the [computationally cheap parameters] are suboptimal in many cases and thus fast convergence can indeed not be guaranteed. Additionally, if the convergence is slow, the low-rank Cholesky factors may grow without adding essential information in subsequent iteration steps." In [30] it was thus suggested to combine the subspace projection idea with the ADI recurrence. The projection is performed onto the space spanned by the columns of the current ADI factor, the idea being motivated by the fact the ADI solution factor belongs to the rational Krylov subspace with the same shifts as ADI. The projection is performed every $\hat{k}$ ADI iterations by computing and orthonormal basis spanning the range of the current factor, and the small size projected equation is solved by means of a Schur-type method (cf. section 5.1). Since the number of columns grows at each iteration, the cost of computing the orthonormal basis significantly increases. To overcome this problem, the authors suggest to truncate the obtained projected solution so that a small rank factor is retained for the next ADI iteration. More technical implementation details can be found in [30]. The idea is very reminiscent of a restarting process in the standard linear system framework, although here the truncation is performed in a different fashion. To complete this parallel with linear system solves, this procedure may be viewed as a hybrid restarted process, where a rational function (here the ADI single step) is applied to the solution before restart; cf., e.g., [195] for polynomial acceleration procedures of restarted methods in the linear system setting. The resulting process is called the Galerkin projection accelerated LRCF-ADI (LRCF-ADI-GP). Note that although ADI does not require that A be either positive or negative definite (that is its field of values in a half complex plane), the extra projection step is ensured not to break down only under the additional definiteness constraint. It is also interesting to observe that, without the truncation of the projected solution, the procedure might be mathematically equivalent to the Galerkin method in the Rational Krylov subspace obtained with the same shift parameters; a formal proof still needs to be carried out.

We also mention the procedure proposed in [143], where the continuous Lyapunov

equation is first transformed into a discrete (Smith) Lyapunov equation with rational matrix functions as coefficient matrices, and then solved by means of the Global Krylov subspace method. This may be viewed as a preconditioning strategy.

5.2.3. Spectral, sparse format and other methods. As for the Sylvester equation, the Kronecker formulation can be used to restate the matrix equation as the following very large linear system,

$$\mathcal{A}\mathbf{x} := (I_m \otimes A + A^\top \otimes I_n)\mathbf{x} = c, \qquad \mathbf{x} = \mathrm{vec}(\mathbf{X}),\ c = \mathrm{vec}(C) \tag{5.10}$$

of size n^2, if n is the size of A; see, e.g., [127] for an early attempt. For A symmetric and positive definite, the convergence rate of CG applied to the Kronecker formulation is driven by the condition number $\kappa(\mathcal{A}) = \kappa(A)$, whereas the convergence rate of the Galerkin procedure directly applied to the original Lyapunov equation is driven by $\kappa(A + \lambda_{\min} I)$ (cf. Theorem 5.3), which can be significantly smaller than $\kappa(A)$. This analysis justifies the better performance of projection methods applied to the matrix equation. A second possibly stronger argument is given by memory requirements: the Kronecker formulation requires n^2-length vectors. Nonetheless, it was recently shown in [190] that when solving (5.10) floating point operations can be carried out so as to lower memory storage from $\mathcal{O}(n^2)$ to $\mathcal{O}(n)$. Moreover, a standard Krylov subspace method for (5.10) can take full advantage of the structure, since most matrix-vector multiplications can be rewritten as matrix-matrix operations.

A possible way to overcome slow convergence is to choose an effective preconditioning strategy, that can drastically improve the spectral properties of the coefficient matrix $\mathcal{A}$. Hochbruck and Starke used a Krylov subspace solver for the system (5.10), and they investigated SSOR and ADI iteration (with a fixed number of iterations) as operator-based preconditioners; see also [190] for some implementation aspects of preconditioning strategies. More recently, a flexible GMRES approach was proposed in [46], which allowed for a variable ADI preconditioning step. Very preliminary numerical results report promising performance of the Kronecker formulation, while taking into account the matrix structure. These approaches may have broader applications for more general matrix equations; see the discussion in section 7.2.

A rather different approach consists of using an appropriately modified version of the sign function iteration depicted in (5.3). As memory requirements are excessive in its original form for large scale problems, two major amendments have been explored (see, e.g., [16]): i) A sparsified version of A, so as to substantially reduce the computation and storage of Z_k^{-1}; ii) For $C = C_1 C_1^\top$, a factored version of the approximation $\widetilde{\mathbf{X}}$, so that only a full (low rank) factor need be iterated. The latter problem was addressed in [29], where the following coupled iteration was proposed:

$$A_0 = A, \quad B_0 = C_1, A_{k+1} = \frac{1}{2}(A_k + A_k^{-1}), \quad B_{k+1} = \frac{1}{\sqrt{2}}[B_k,\ A_k^{-1}B_k],\ k = 0, 1, \ldots,$$

giving $\mathbf{Y} = \frac{1}{\sqrt{2}} \lim_{k\to\infty} B_k$, with $\mathbf{Y}\mathbf{Y}^\top = \mathbf{X}$. Note that the size of B_k is doubled at each iteration, therefore a rank reduction is suggested in [29]. More recent developments include new algorithms that appear to be well suited for large problems with dense and unstructured matrices [230]; the discussion in [230] in fact addresses generalized Sylvester equation.

Item i), namely reducing the cost of dealing with the explicit inverse of large matrices, may be addressed by exploiting data-sparse matrix representation and approximate arithmetic. In [18], but also in previous related works for the algebraic

Riccati equation, the $\mathcal{H}$-matrix format was used (cf. section 4.4.3). If $\mathrm{Inv}_{\mathcal{H}}(A)$ denotes the inverse in the $\mathcal{H}$-matrix format, then the coupled recurrence above can be performed as

$$A_{k+1} = \frac{1}{2}(A_k + \mathrm{Inv}_{\mathcal{H}}(A_k)), \quad B_{k+1} = \frac{1}{\sqrt{2}}[B_k,\ \mathrm{Inv}_{\mathcal{H}}(A_k)B_k],\ k = 0, 1, \ldots,$$

where the sum to obtain A_{k+1} is intended in $\mathcal{H}$-matrix format. More implementation details can be found in [18]. According to the analysis performed in there, the error induced by the new format can be controlled while performing the rank reduction of B_{k+1}, so that the format errors do not grow unboundedly with k; these results are in agreement with the general theory of $\mathcal{H}$-matrices for Riccati equations, developed in [105]. In [18], the derivation with $\mathcal{H}$-matrix format is extended to the case of the generalized Lyapunov equation (cf. section 7). Numerical experiments show that the $\mathcal{H}$-format allows the sign function iteration to be employed for medium size problems ($\mathcal{O}(10000)$), for which the dense algorithm requires too much memory allocation. Finally, an example comparing a linear multigrid solver using $\mathcal{H}$-format matrices with ADI is reported in [103, sec. 7.6], showing that on that specific example the multigrid approach is about ten times faster than ADI (implemented in the same framework), although both methods scale linearly with the number of multigrid levels.

We conclude this section with strategies that are more explicitly based on invariant subspaces. All considered methods assume that the maximum rank of a sufficiently accurate approximate solution is either known or given.

The integral representation of $\mathbf{X}$ in (4.5) and the spectral decay of the spectrum of $\mathbf{X}$ suggest various eigenvalue-based strategies. One such method focuses on approximating the leading invariant subspace of $\mathbf{X}$. In [131] and [128] an Approximate Power Iteration (API) approach was proposed, which aims to approximate the dominant eigenvectors of $\mathbf{X}$. The method is closely related to power iteration and the Lanczos method for computing the extremal eigenpairs of a positive definite symmetric matrix, and the authors report good convergence properties when the eigenvalues associated with the sought after eigenvectors are away from zero and well separated from the others, so that a good low-rank approximation of $\mathbf{X}$ can be determined. The method works under the assumption that A is negative definite, as with projection methods. The API method applies the power method to $\mathbf{X}$, which is only known implicitly and approximately by means of products of the type $y = \mathbf{X}v$ through the solution of an auxiliary Sylvester equation (cf. section 4.3). The numerical experiments reported in [128] on actually quite small problems, seem to imply that API is a promising method for the approximation of the leading eigenvectors of $\mathbf{X}$, without the computation of $\mathbf{X}$ itself. The approach is reminiscent of the implicitly restarted Arnoldi method [173], although each iteration requires the solution of a Sylvester equation. A variant of this approach was proposed in [264] to overcome misconvergence caused by the omission of a term in the error matrix. Motivated by [128], an algorithm combining the power method and (implicitly) the ADI iteration was proposed in [197]; we refer to [196] for a more thorough presentation of these approaches.

With the same aim of approximating the leading invariant subspace of $\mathbf{X}$ of given dimension, the procedure explored in [110] performs a refined numerical approximation by repeatedly integrating the dynamical system associated with the Lyapunov equation, as the basis for an orthogonal power iteration.

A somewhat related approach was proposed in [237], and it exploits the popular *proper orthogonal decomposition* (POD) approach employed in reduced order modeling

of large-scale dynamical systems [28]. The idea is to collect a sample of m solutions to the associated linear differential equation with different starting data, and for a chosen k, form a rank-k approximate Lyapunov solution by using the eigenpairs of the sample, obtained by the method of snapshots. The approach relies on the integral representation of the Lyapunov solution, and according to the author, it is particularly appropriate for infinite dimensional problems.

Finally, a novel very different approach was recently proposed by Vandereycken and Vandewalle in [261] for A symmetric and positive definite; the method finds a low-rank approximation to $\mathbf{X}$ by minimizing the function

$$f : \mathcal{M}_k \to \mathbb{R}, \quad X \mapsto \operatorname{trace}(XAX) - \operatorname{trace}(XC)$$

on the manifold of symmetric and positive semidefinite matrices of rank k in $\mathbb{R}^{n\times n}$, namely

$$\min_{X\in\mathcal{M}_k} f(X).$$

Note that the minimization of f corresponds to the minimization of the error in the energy norm, having used the Kronecker formulation of the Lyapunov equation. By using the smoothness of $\mathcal{M}_k$ the problem is solved within a Riemann optimization framework, which allows one to embed the rank constraint in the space, and solve an unconstrained minimization problem by means of a Riemann trust-region method, a second-order model based on the Hessian [2]. At convergence of the minimization process, if the current solution rank is not sufficiently accurate, the process is restarted basically from scratch. As a result, the method may be appealing when the optimal rank is approximately known a-priori, otherwise the approach may not be competitive with respect to other strategies discussed so far.

6. The Stein and discrete Lyapunov equations. The Stein and the discrete Sylvester equations are the discrete-time counter part of the (continuous-time) equations discussed in the previous sections, and they naturally stem from a discrete-time system; see, e.g., [4, sec.4.3]. Other relevant applications include for instance statistics [155],[154], probability [10], spectral analysis [134]; these equations are also a computational tool in the design of control systems [120], or in the coprime matrix fraction description of linear systems [290].

The Stein equation may be written as

$$\mathbf{X} + A\mathbf{X}B = C, \tag{6.1}$$

where it is assumed that the eigenvalues of A and B are contained in the open unit disk. The discrete-time Lyapunov equation is obtained by choosing $B = -A^\top$, in which case, if a solution exists, it has to be symmetric. In the context of inertia theory, for $C \geq 0$ the discrete-time Lyapunov equation allows one to analyze the connection between the eigenvalues location of the matrix A with respect to the unit circle and the eigenvalue location of $\mathbf{X}$ with respect to the imaginary axis; see, [169, sec.13.2], and also, e.g., [279],[174] for more specialized results.

Under the condition that $\lambda_i(A)\lambda_j(B) \neq -1$ for all i, j, the solution $\mathbf{X}$ exists and is unique for any C (see, e.g., [168]), and this is highlighted by the Kronecker form of (6.1), given as $(I + B^\top \otimes A)\mathbf{x} = c$, where $\mathbf{x} = \operatorname{vec}(\mathbf{X})$ and $c = \operatorname{vec}(C)$. Necessary and sufficient conditions for the existence and uniqueness of the solution $\mathbf{X}$ were obtained in [278] as a generalization of the property (4.2) for the Sylvester equation. Inertia

and other transformation-based results for $B = -A^\top$ can be derived in a natural manner from those on the Lyapunov equation; see, e.g., [227],[228]. We also refer to [167] for a solution expressed in terms of the companion form of the given matrices, and [40] for related computational considerations.

For numerically solving the equation for, say, B nonsingular, one could think of working with $\mathbf{X}B^{-1} + A\mathbf{X} = CB^{-1}$ which is a continuous-time Sylvester equation, and then adopt one of the solution methods from previous sections. In fact, (6.1) is nothing but a generalized Sylvester equation as in (7.1) with special choices of the first two coefficient matrices. For large B, the matrix B^{-1} should not be formed explicitly, but its action used within iterative methods.

Forming B^{-1} explicitly is not recommended also in the small size case, whenever B is ill-conditioned. Alternative transformations that bring the discrete equation to standard form are given by (here for $B = -A^\top$, cf. [208])

$$\tilde{A}\tilde{\mathbf{X}} + \tilde{\mathbf{X}}\tilde{A}^\top = C, \quad \text{with } \tilde{A} = (A - I)^{-1}(A + I),\ \mathbf{X} = \frac{1}{2}(\tilde{A} - I)^\top \tilde{\mathbf{X}}(\tilde{A} - I),$$

and (cf. [12], [206], [152])

$$\tilde{A}\mathbf{X} + \mathbf{X}\tilde{A}^\top = \tilde{C}, \quad \text{with } \tilde{A} = (A - I)(A + I)^{-1}, \tilde{C} = 2(A^\top + I)^{-1}C(A + I)^{-1},$$

where it is assumed that the inversions are well defined. In general, however, the same stability considerations as for methods using A^{-1} apply.

All these difficulties encourage solving the discrete equations (6.1) directly. A Schur-form type method for small size coefficient matrices that directly deals with (6.1) can be found, e.g., in [13], while a generalization of the "continuous-time" Hammarling method was proposed by Hammarling himself in [118].

In [263], Varga established a rank-two updating formula for the Cholesky factors in Hammarling's algorithm for solving the real, non-negative definite Stein equation. As a generalization of his algorithm for the Lyapunov equation, a block variant of the Hammarling's method for the discrete-time Lyapunov equation is suggested in [161]. We also mention a systolic algorithm with coefficient matrices in triangular form [135].

In spite of the large similarity with the standard equation, directly attacking (6.1) is an interesting problem in its own, especially for A and B of large dimensions and with either of the two matrices singular. For a low rank C, projection methods are applicable to solve (6.1), and an approximate solution $\widetilde{\mathbf{X}} = V_k \mathbf{Y} W_k^\top$ can be determined, where the columns of V_k and W_k span approximation spaces associated with A and $B^\top$, respectively. For instance, a global Krylov subspace approach was proposed in [142, sec. 5],[148], and its implementation is a natural modification of that used for the standard Sylvester equation. Similar derivations can be obtained for other Krylov-based methods. For instance, the Extended Krylov subspace could be easily adapted to the transformed matrix equation for low rank C, at the same computational cost as for the standard Sylvester equation.

The discrete-time Lyapunov equation motivated the development of the Smith method [239], which is at the basis of the modern ADI iteration for the Lyapunov equation. For A d-stable, the unique solution to (6.1) can be written as $\mathbf{X} = \sum_{j=0}^{\infty} A^j C (A^j)^\top$, and it is symmetric and positive semi-definite, if C is. The (convergent) Smith iteration is defined as

$$\mathbf{X}_0 = 0, \quad \mathbf{X}_{k+1} = C + A\mathbf{X}_k A^\top,$$

with a closed form given by $\mathbf{X}_k = \sum_{j=1}^{k} A^{j-1} C (A^{j-1})^\top$. Faster - quadratic - convergence can be achieved with the *squared* Smith method, which becomes of interest in the large scale case precisely for C of small rank [203]. The iteration is generically given as

$$\mathbf{X} = (A^{2^{k+1}})^\top \mathbf{X} A^{2^{k+1}} + \sum_{i=0}^{2^{k+1}-1} (A^i)^\top H A^i, \quad \mathbf{X} = -\lim_{k\to\infty} \sum_{i=0}^{2^{k+1}-1} (A^i)^\top C A^i.$$

The resulting recursion is given by $H_{k+1} = H_k + A_k^\top H_k A_k$ where $A_{k+1} = A_k^2$, so that $C_k \to \mathbf{X}$ as $k \to \infty$. By exploiting the low rank of $C = C_0 C_0^\top$, $H_{k+1} = C_{k+1} C_{k+1}^\top$ with $C_{k+1} = [C_k, A_k^\top C_k]$. As a result, the number of columns of C_{k+1} doubles at each iteration, and C_{k+1} is contained in a block Krylov subspace generated by $A^\top$ and C_0. Recent advances to make this recurrence more effective both in terms of computational costs and memory requirements include compressions, truncations and restarts, with a tricky use of the underlying Krylov subspace [178],[177],[222],[33]. In these references, estimates for the residual and error norms are also derived. Finally, we point out an ADI acceleration strategy in [222] (for $B = -A^\top$) and in [33], which significantly improves the convergence speed. In fact, a major breakthrough for the Smith method consisted in combining its recurrence with the ADI idea, as developed in [203].

All these approaches rely on the fact that often the solution $\mathbf{X}$ has (numerical) rank much lower than n; indeed, in [33] it is shown for the Stein equation that if C has rank p,

$$\frac{\sigma_{kp+1}(\mathbf{X})}{\sigma_1(\mathbf{X})} \le \|A^k\| \|B^k\|,$$

indicating that indeed the solution rank may be small. In [222] the following estimate was derived for $B = -A^\top$ and $\|A\| < 1$:

$$\frac{\sigma_{kp+1}(\mathbf{X})}{\sigma_1(\mathbf{X})} \le \frac{\|A\|^{2k}}{1 - \|A\|^2}.$$

Other variants of the Smith iteration have been considered. For instance, in [291], for $r \ge 2$ an iteration of the type

$$X_{k+1} = \sum_{j=0}^{r-1} A_k^j X_k B_k^j, \quad \text{with} \quad A_{k+1} = A_k^r,\ B_{k+1} = B_k^r,\ k \ge 0$$

was introduced. The performance of the method is yet to be tested. In general, a computational comparison of all these approaches and variants is still lacking, though it would be highly desirable.

A related matrix equation is the $\top$-Stein equation, given by $\mathbf{X} = A\mathbf{X}^\top B + C$, whose solvability conditions have been recently analyzed in [180]. More generally, a broader class of matrix equations can be written as $\mathbf{X} = Af(\mathbf{X})B + C$, where $f(X) = X^\top$, $f(X) = \bar{X}$ or $f(X) = X^*$, whose analysis and numerical solution can be recast in terms of the Stein matrix equation [292]. This and more general forms of linear equations are discussed in the next section.

7. Generalized linear equations.

7.1. The Generalized Sylvester and Lyapunov equations. The term *generalized* refers to a very wide class of equations, which includes systems of matrix equations, bilinear equations and problems where the coefficient matrices are rectangular. We start with the most common form of generalized Sylvester equation, namely

$$A\mathbf{X}D + E\mathbf{X}B = C, \tag{7.1}$$

which differs from (4.1) for the occurrence of coefficient matrices on both sides of the unknown solution $\mathbf{X}$.

If D and E are both nonsingular, left multiplication by E^{-1} and right multiplication by D^{-1} lead to a standard Sylvester equation, with the same solution matrix $\mathbf{X}$. In case either E or D are ill-conditioned, such a transformation may lead to severe instabilities. This problem is common to other *generalized* equations we will encounter later in this section, and it justifies the development of solution methods that stick to the original form (7.1). The case of singular D and/or E is called *projected* Sylvester equation, and it has an important role, for instance, in the solution of differential-algebraic equations [166]; as already mentioned, we shall not discuss this setting in this work.

The following result ensures existence of a unique solution $\mathbf{X}$ to (7.1).

THEOREM 7.1. ([267]) *The matrix equation $A\mathbf{X}D + E\mathbf{X}B = C$ has a unique solution if and only if*

(i) The pairs (A, E) and $(D, -B)$ are regular pencils;

(ii) The spectra of (A, E) and $(B, -D)$ are disjoint[9]*.*

We also refer to [201] for some estimator of the separation[10] and the condition number of the operator associated with (7.2), which is important to assess the accuracy of the computed solution. For C positive semidefinite and (A, E) stable, in [201] a generalization of the Hammarling method is also proposed.

A natural extension of the Bartels-Stewart method can be implemented for numerically solving (7.2) when dimensions are small, and this was discussed in [94], [95], [201], where the starting point is a QZ decomposition of the pencil (A, E), followed by the solution of a sequence of small (1-by-1 or 2-by-2) generalized Sylvester equations, which is performed by using their Kronecker form.

The large scale setting does not significantly differ from previous cases, as long as E, D are not too ill-conditioned. The problem can be recast as a standard Sylvester equation in $E^{-1}A$ and BD^{-1}. In the case of rational Krylov subspace and ADI methods, shifted systems can be solved with the coefficient matrix $(E^{-1}A + sI) = E^{-1}(A + sE)$, and analogously for systems with BD^{-1}.

We mention the specific application of global Krylov subspace methods (cf. section 4.4), which are obtained by using the mapping $\mathcal{M}(X) = AXD + EXB$, therefore they can be applied in general to the equation $\sum_{i=1}^{q} A_i X B_i = C$, as done in [48]; note that this kind of approach can only be applied to medium size problems, as the matrix formulation involves dense matrices. As a related issue, we recall once again that there is a tight relation between global methods and the Kronecker form, providing a good ground for the theoretical understanding of the performance of global methods.

[9] Here the notion of disjoint spectra ([267, formula (7)]) should be intended keeping in mind the definition of "spectral set" in generalized eigenvalue problems, as defined for instance in [246, Definition VI.1.1].

[10] Defined as $\mathrm{sep}_p(A, E) = \min_{\|X\|_p=1} \|A^\top XE + E^\top XA\|_p$, with $p = 2, F$.

As particular case of the linear equation in (7.2), the generalized Lyapunov equation

$$A\mathbf{X}E^\top + E\mathbf{X}A^\top = C, \tag{7.2}$$

has a special interest in control; see also more recent applications in stability analysis [189]. The case $E \neq I$ arises in a control problem, for instance, whenever a second or higher order ordinary differential equation is discretized. A unique solution is ensured if and only if Theorem 7.1 applies, that is all eigenvalues of the pencil (A, E) are finite, and they do not have pairwise zero sum. As a consequence, a unique solution is only obtained if one of the matrices A, E is nonsingular. In this case one can recast equation (7.2) as a standard Lyapunov equation.

To avoid stability problems caused by a possible ill-conditioned E or A, it is usually preferred to work with E and A implicitly. This is realized by performing a simultaneous Schur decomposition of E and A, $E = QSZ^*$, $A = QTZ^*$, with S and T (complex) upper triangular [191]. Plugging in this transformation, (5.2) becomes $QTZ^*\mathbf{X}ZS^*Q^* + QSZ^*\mathbf{X}Z^*T^*Q^* = C$, that is

$$T\widehat{\mathbf{X}}S^* + S\widehat{\mathbf{X}}T^* = Q^*CQ, \quad \widehat{\mathbf{X}} = Z^*\mathbf{X}Z.$$

The elements of $\widehat{\mathbf{X}}$ can then be obtained by exploiting the structure of T and S [117].

A different approach adapts the matrix sign function iteration in (5.4) to this more general context, and it is shown in [29] that it is applicable under the hypothesis that the Lyapunov equation is stable, that is that the pencil (A, E) has all eigenvalues in the open left complex half-plane. In the case of C in factorized form in (7.2), a recurrence is proposed in [29] to generate an approximation to the Cholesky-type factor of the resulting semidefinite solution $\mathbf{X}$. Comparisons in terms of memory requirements and floating point operations with respect to the generalized Hammarling method (see [201]) are also reported in [29].

7.2. Bilinear, constrained, and other linear equations. Other generalizations of the Sylvester equation have attracted the attention of many researchers. In some cases the standard procedure for their solution consists in solving a (sequence of) related standard Sylvester equation(s), so that the computational core is the numerical solution of the latter by means of some of the procedures discussed in previous sections. We thus list here some of the possible generalizations more often encountered and employed in real applications. We start by considering the case when the two coefficient matrices can be rectangular. This gives the following equation:

$$A\mathbf{X} + \mathbf{Y}B = C, \tag{7.3}$$

where $\mathbf{X}$, $\mathbf{Y}$ are both unknown, and A, B and C are all rectangular matrices of conforming dimensions. Equations of this type arise in control theory, for instance in output regulation with internal stability, where the matrices are in fact polynomial matrices (see, e.g., [277] and references therein). The following theorem is a first result on the existence and uniqueness of the pair $\mathbf{X}$, $\mathbf{Y}$, and is reported as originally stated in [217]; see also more recent advanced developments in [85].

THEOREM 7.2. ([217]) *The necessary and sufficient condition that the equation $A\mathbf{X} - \mathbf{Y}B = C$, where A, B, and C are $m \times r$, $s \times n$, and $m \times n$ matrices respectively with elements in a field $\mathcal{F}$, have a solution $\mathbf{X}, \mathbf{Y}$ of order $r \times n$ and $m \times s$ respectively and with elements in $\mathcal{F}$ is that the matrices*

$$\begin{bmatrix} A & C \\ 0 & B \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix}$$

be equivalent.

These are extensions of the analogous result for the standard Sylvester equation, cf. (4.2) and [217], [132]. The two-sided version of (7.3) is given by

$$A\mathbf{X}D + E\mathbf{Y}B = C,$$

and this is an example of more complex bilinear equations with several left-hand side terms considered in the literature; see, e.g., [290] and references therein. A typical setting is given by the following bilinear equation:

$$A\mathbf{X}D + E\mathbf{X}B = C\mathbf{Y} + F, \tag{7.4}$$

where the pair $(\mathbf{X}, \mathbf{Y})$ is to be determined, and $\mathbf{X}$ occurs in two different terms. Theoretical aspects are collected in [282], and also in [281], [283], where closed forms for $(\mathbf{X}, \mathbf{Y})$ in the homogeneous ($F = 0$) and non-homogeneous cases are developed, respectively. We refer to [288] for interesting explicit equivalence relations among the homogeneous form of (7.4), and other linear generalized Sylvester equations; the transformations employed are similar to those we have described in section 6 to transform discrete-time equations into continuous-time equations. In [287],[289] general parametric expressions for the solution matrices $\mathbf{X}$ and $\mathbf{Y}$ are also obtained, under the hypothesis that D is full rank and F is the zero matrix.

The main objective in the aforementioned papers is in fact the solution of *systems* of bilinear matrix equations; see e.g. [151],[81], for which a recent perturbation analysis can be found in [183]. These systems can arise, for instance, in the numerical treatment of systems of stochastic partial differential equations; currently, a Kronecker formulation is used to determine the numerical solution of the resulting large linear system; see, e.g., [87]. Systems of linear matrix equations also arise in the spectral computation of pencils [267], and we refer to [150] for solution methods for small size problems, and for further discussion on the solvability conditions and applicability of these systems. In [209] a necessary and sufficient condition is given for the matrix equations

$$\begin{cases} A_1\mathbf{X} + \mathbf{Y}B_1 = C_1 \\ A_2\mathbf{X} + \mathbf{Y}B_2 = C_2 \end{cases} \tag{7.5}$$

to have a pair of common solutions $\mathbf{X}$ and $\mathbf{Y}$. From a numerical standpoint, very few alternatives have been explored so far, that go beyond a cleverly implemented Kronecker formulation. The idea suggested in [81] amounts to "expanding" the two equations into a single one of larger size, whose solution contains both $\mathbf{X}$ and $\mathbf{Y}$ but requires the Jordan decomposition of some of the coefficient matrices; necessary and sufficient conditions for determining such a unique solution are also obtained. A similar framework is used in [74] where more than two unknown matrices are allowed, and an approximate solution is obtained by means of a least squares approach. In [151], a more stable generalized Schur method is proposed, which applies the QZ algorithm to the pairs of coefficient matrices.

The number of linear matrix equations and unknown matrices can in fact be quite large, as discussed for instance in [47]. Necessary and sufficient conditions for the resulting systems to have a solution pair are studied in [273]. Computationally speaking, this general case has only been treated so far by using the Kronecker formulation, so that only very small problems have been tackled; but see [290], where the problem of solving the set of matrix equations is recast as an optimization problem.

A special class of *nonlinear* problems is given by the following Sylvester-Observer equation, which stems from the problem of determining a *reduced* order observer model [75], [185]. Find matrices $\mathbf{X}, \mathbf{Y}$ and $\mathbf{Z}$ such that

$$\mathbf{X}A - \mathbf{Y}\mathbf{X} = \mathbf{Z}C, \quad \begin{bmatrix}\mathbf{X}\\ C\end{bmatrix} \text{ invertible}, \tag{7.6}$$

where A and C are known matrices with C having few rows. A solution to (7.6) exists for any choice of spectrum of $\mathbf{Y}$, therefore this spectrum can be predetermined; a choice that makes $\mathbf{Y}$ stable (eigenvalues in the left half complex plane) also ensures convergence of the reduced order observer; we refer to [185] for more details on these aspects. A possible way to address (7.6) is to choose $\mathbf{Y}$ and $\mathbf{Z}$ arbitrarily and then solve for $\mathbf{X}$ the resulting Sylvester equation. Early approaches in this direction did not lead to a numerically stable method. For small size matrices, the reduction to Hessenberg form proposed by Van Dooren in [75] is still one of the most effective methods for solving (7.6). The algorithm is based on a reduction to "staircase form" of the pair (A, C), and on the determination of the solution $\mathbf{X}$ with a particular structure in a different coordinate system. We also refer to [259] for a more detailed survey on methods for dense matrices. More recently, other approaches have been proposed: for instance, a block generalization of the method in [75] was proposed in [56]; moreover, in [57] the authors proposed a block algorithm for determining a full rank solution, and it seems to be most appropriate for large-scale problems with sparse A. In this latter setting, a successful early method was proposed in [67]. The approach first assumes that $\mathbf{Z}C$ is rank-one and exploits the resemblance between the observer-Sylvester equation and the Arnoldi relation (4.16). As a by-product of the method, the authors in [67] also derive an algorithm for solving the partial-pole-assignment problem for large and sparse A, which is generalized in [69] to higher rank of $\mathbf{Z}C$; see also [71] for alternative algorithmic approaches. The authors in [54] propose a new strategy for a-priori choosing the eigenvalues of $\mathbf{Y}$ that makes the algorithm in [67] more efficient. From a control theory point of view, the possibility to determine a reduced order model is also important in the derivation of stable closed-loop systems, giving rise to a well exercised eigenvalue assignment problem. We refer to, e.g., [70], for a brief survey on this and other related problems.

Within the Sylvester-observer equation, we can formulate the problem in a slightly different manner, namely by means of a constraint (see, e.g., [248],[256],[192]), and it can be stated as follows (cf., e.g., [11]): *Given* $A \in \mathbb{R}^{n\times n}$ *stable,* $B \in \mathbb{R}^{n\times p}$, $C \in \mathbb{R}^{m\times n}$ *and* $F \in \mathbb{R}^{(n-m)\times(n-m)}$, *find* $\mathbf{Y} \in \mathbb{R}^{(n-m)\times m}$ *and* $\mathbf{X} \in \mathbb{R}^{(n-m)\times n}$ *such that*

$$\mathbf{X}A - F\mathbf{X} = \mathbf{Y}C, \quad \text{with} \quad \mathbf{X}B = 0, \tag{7.7}$$

so that $[\mathbf{X}; C]$ *is full rank.*

The problem may be viewed as a homogeneous system of linear matrix equations, generalization of (7.5), with two terms in $\mathbf{X}$ as in (7.4), and $C_1 = 0$, $C_2 = 0$; however, there is no need to expand it by means of the Kronecker product. In [11] the authors provide necessary and sufficient conditions for a solution to (7.7) to exist, and propose an algorithm for its computation in the small-scale case. The main ingredients are a QR factorization and the solution of a standard Sylvester equation. In [96] a modification of this method was presented, to handle the case of almost singular Sylvester equations. The large scale case has been recently addressed in [226] by generalizing the method in [11]. We also refer to [58] for an approach that handles a "regional pole-placement constraint" on F in (7.7) for a descriptor system, and to [66] and its

references for further theoretical properties. Somewhat similar difficulties arise when dealing with projected Lyapunov equations, as discussed in [182].

Going back to a single unknown matrix, other "generalized" Lyapunov equations more in the spirit of (1.2) include extra linear terms:

$$A\mathbf{X} + \mathbf{X}A^* + N\mathbf{X}N^* + C_1C_1^* = 0, \tag{7.8}$$

and they stem, for instance, from complex systems of the form:

$$\dot{x}(t) = (A + iu(t)N)x(t) + iB_1u(t), \quad x(0) = x_0, \quad y(t) = C_1^*x(t),$$

see, e.g., [119],[223]. We refer to [107] for sufficient conditions on the existence of the controllability and observability Gramians; more complex forms involve more structured matrices N, see, e.g., [24],[62]. In fact, more terms of the type $N_j\mathbf{X}N_j^*$, $j = 1, 2, \ldots$ could arise to fulfill more general model requests. The one above is an example of linear jump systems (see [187]), in which the linear coefficient matrices depend on a Markovian random process, giving rise to systems of matrix equations with the extra term, accounting for the probabilistic nature of the problem. Polynomial and infinite dimension systems are also of interest, see, e.g., [284] and [64], respectively, and their references. In addition to a robust Kronecker-form based iteration reviewed in [65, sec.3.1-4], Damm in [65] proposed a regular splitting for the numerical solution of (7.8), yielding the following iterative scheme:

$$A\mathbf{X}_{k+1} + \mathbf{X}_{k+1}A^* = -N\mathbf{X}_kN^* - C_1C_1^*, \quad \mathbf{X}_0 = 0,$$

which entails the solution of a sequence of standard Lyapunov equations. Convergence to $\mathbf{X}$ is obtained if the spectrum of A is sufficiently away from the imaginary axis. We refer to [65, sec. 4], where the generalized case of the Lyapunov operator is also treated. In the recent article [32] a thorough discussion and contextualization of the algebraic problem in stochastic model order reduction can be found. In [24], various methods for the Lyapunov equation, such as ADI and projection techniques, are adapted to the setting of (7.8), including sparse format approaches for the Kronecker formulation; reported experimental results on large problems seem to favor this latter approach, with the caveat of tuning the sparsity and accuracy parameters, as described in section 5.2.3.

An approach that may be appropriate for large scale problems is implicitly suggested in [8]. In the context of model order reduction, the following approximation space is introduced:

$$\text{range}(V) = \text{range}\left\{\bigcup_{k=1}^{r} \text{range}\{V^{(k)}\}\right\}, \tag{7.9}$$

with $\text{range}(V^{(1)}) := K_q(A^{-1}, A^{-1}C_1)$ and

$$\text{range}(V^{(k)}) := K_q(A^{-1}, A^{-1}NV^{(k-1)}), k = 2, \ldots, r.$$

Using a Galerkin approximation onto range(V), the equation (7.8) can be reduced and solved with a direct procedure; a possible implementation of this idea was recently performed in [31]. Another approach for solving multilinear systems in Kronecker form was analyzed in [165], for which a tensor-based form for the approximate solution is considered. Such strategy is well suited in the approximation of *parameterized* linear

systems, as they arise for instance in certain discretization strategies for the numerical solution of stochastic partial differential equations [6]. Data sparse methods associated with the Kronecker formulation may provide a possible successful venue for attacking the general linear multi-term matrix equation (1.2); to the best of our knowledge, no attempts have been made in this direction so far for really large problems.

7.3. Sylvester-like and Lyapunov-like equations. Sylvester and Lyapunov-like linear matrix equations of the form ([49], [41])

$$B\mathbf{X} + f(\mathbf{X})A = C, \quad A^*\mathbf{X} + f(\mathbf{X})A = C, \quad B, A, \mathbf{X} \in \mathbb{C}^{m\times n}, \tag{7.10}$$

with $f(X) = \bar{X}$, $f(X) = X^\top$, $f(X) = X^*$, or their "discrete-time" variants (cf. section 6) are less common; but see, for instance, [164] for an occurrence in structured eigenvalue computation. The homogeneous case ($C = 0$) has been recently analyzed in [253], where a complete description of the solution in terms of the Kronecker canonical form of $A + \lambda f(B)$ is derived, whenever information on this latter pencil is known. These equations have attracted increasing interest in the past few years, with recent contributions on the necessary and sufficient conditions for the solvability of this equation, for any right-hand side matrix C [136]; a different proof of this result that also induces a numerical method is proposed in [266]. As a sample of this type of result, in [53, Lemma 5.10] for $f(X) = X^\top$, it is proved that a unique solution $\mathbf{X}$ exists if and only if the pencil $A - \lambda B^\top$ is regular and if its spectrum is $\top$-reciprocal free, with possibly the only exception of the unit eigenvalue, which should be simple.

In [251], an algorithm that relies on the generalized Schur decomposition of the pair $(A, f(B))$ (via the QZ algorithm) is proposed to determine $\mathbf{X}$ for small A and B. For $f(X) = X^\top$ this can be briefly summarized as follows:

1) Decompose $A = URV$ and $B^\top = USV$, with U, V unitary and R, S upper triangular;

2) Compute $E = VCV^\top$;

3) Solve $S^\top W + W^\top R = E$ element-by-element;

4) Form $X = \bar{U}W\bar{V}$.

The solution of the equation in step 3) is also treated in detail in [251]. The numerical solution in the large scale case is currently an open problem.

In [285] a closed-form solution to the equation for $f(X) = \bar{X}$ is considered, together with the set of all possible solutions for (7.4) and for the bilinear problem $A\bar{\mathbf{X}} + B\mathbf{Y} = \mathbf{X}F$. Chiang, Duan, Feng, Wu and their collaborators have thoroughly investigated these formulations and their role in control applications.

A particular extension of this class of problems is given by polynomial equations. Consider a polynomial matrix $R(\xi) = R_0 + R_1\xi + \cdots + R_\ell\xi^\ell$ in the unknown ξ, where R_i are constant square matrices and such that $\det(R(\xi))$ is not identically zero, and let Z be a square polynomial matrix satisfying $Z(\xi) = Z(-\xi)^\top$. The equation

$$R(-\xi)^\top\mathbf{X}(\xi) + \mathbf{X}(-\xi)^\top R(\xi) = Z(\xi),$$

in the square polynomial matrix $\mathbf{X}$ is called the polynomial Sylvester equation. This special equation plays a role in the computation of integrals of quadratic functions of the variables of a system and their derivatives (see [249],[250]), and in the stability theory for high-order differential equations. In [200] the authors focus on the case when the right-hand side has the form $Z(\xi) = Q(-\xi)^\top\Sigma Q(\xi)$, where Q is a real rectangular polynomial matrix in ξ such that QR^{-1} is a matrix of strictly proper rational functions and Σ is a diagonal, signature matrix. An iterative solution method

inspired by the method of Faddeev for the computation of the matrix resolvents is also derived. More general equations include polynomial Diophantine matrix equations in the form $D(\xi)\mathbf{X}(\xi) + N(\xi)\mathbf{Y}(\xi) = F(\xi)$; in [120] closed form solutions are presented, which could be used to numerically solve small size equations. In the large scale setting, this problem is computationally unresolved.

Finally, a special mention deserves the homogeneous version of the Sylvester-like equation previously discussed with $B = A$,

$$A\mathbf{X}^\top + \mathbf{X}A = 0. \tag{7.11}$$

For each fixed complex matrix A, the solution space to this latter equation is a Lie algebra equipped with Lie bracket $[X, Y] := XY - YX$. We refer to the recent articles [252], [93] and their references for more details.

8. Software and high performance computation. Reliable software for solving matrix equations has been available for long time, due to its fundamental role in control applications; in particular, the SLICE Library was made available already in 1986. Early in the 1990' the SLICOT library (`http://www.slicot.org/`,[257]) replaced SLICE, and since then a large number of additions and improvements have been included; see, e.g., [238], [35]. Most recent versions of Matlab [188] also rely on calls to SLICOT routines within the control-related Toolboxes. SLICOT includes a large variety of codes for model reduction and nonlinear problems on sequential and parallel architectures; as a workhorse, both the Bartels-Stewart algorithm and the Hessenberg-Schur algorithm are implemented. The Bartels-Stewart algorithm for triangular matrices is also included as a standard in LAPACK. Functions solving the Lyapunov equation are also available in other computational environments, such as Mathematica [280]. Related projects have lead to the developments of additional codes. Specifically oriented to linear matrix equations in the Matlab framework, the `lyapack` set of routines developed by Penzl in [205] has been particularly successful as a possible implementation of the ADI method for large scale Lyapunov equations, relying on preprocessing for the computation of quasi-optimal parameters. These routines were included in the NICONET Project repository (`http://www.icm.tu-bs.de/NICONET/`). The MESS set[11] by Saak, Mena and Benner is the successor to the `lyapack` package, with the aim of fully exploiting the capabilities of newer releases of Matlab. In addition, MESS allows for the solution to a larger variety of matrix equations associated with the differential Riccati equation. A rather detailed list of routines for solving control-related matrix equations is provided in the book by Sima [229] and in the more recent book by Datta [68].

A number of benchmark problems have been made available for testing purposes. In addition to those available in the NICONET website, a variety of datasets is available in the Oberwolfach collection, accompanied by a well documented description of the originating application problems [61]; see also the description in [163].

Refined implementations of structured linear equations have been proposed for high performance computations. In particular, the efficient solution of triangular and quasi-triangular Sylvester equations has been discussed in [212], [207]. A high performance library for triangular Sylvester-type matrix equations (continuous and discrete-time) is also available at `http://www8.cs.umu.se/~isak/recsy/`, while a parallel SCALAPACK-style version of this software, called SCASY, is available at `http://www8.cs.umu.se/~granat/scasy.html`. Some of the SLICOT routines are

[11] Available at `http://www.en.mpi-magdeburg.mpg.de/mpcsc/mitarbeiter/saak/Software/mess.php?lang=en`

overloaded in these libraries; see [146],[147],[101] for more information on the implementation on parallel architecture.

In [130] an early parallel algorithm was developed to solve medium size ($0 < n \leq 1000$) Lyapunov problems with a banded and negative definite matrix A; experiments with a shared memory multiprocessor machine (Alliant FX-8) can also be found. The approach is similar in spirit to classical iterative linear system methods such as Jacobi and Gauss-Seidel. More recently, specialized parallel algorithms for Lyapunov, Stein and other generalized matrix equations for different modern architectures have been presented by a number of authors; see, e.g., [211] within the Cray T3E, [37],[38] employing a cluster of PCs, [25] within hybrid CPU-GPU platforms. The use of approaches based either on the square Smith iteration or on iterative techniques for the matrix sign function, as opposed to the Schur decomposition, is key to obtain good parallel performance.

Systems of matrix equations were implemented in a parallel environment in [47] and references therein. A parallel algorithm for the small scale solution to the multi-input Sylvester-observer equation (cf. section 7.2) was proposed in [45], and tested on two shared-memory vector machines.

9. Concluding remarks. The solution of linear matrix equations has always attracted the attention of the engineering and scientific communities. The reliability of efficient core numerical algebra methods has made the solution of these matrix equations increasingly popular in application problem modelling. A good understanding of the theoretical tools and of the variety of numerical methods available for Sylvester-type equations provides a solid ground for attacking more general - nonlinear, multiterm or multifunctional - matrix equations, as the recent developments described in previous sections seem to indicate.

We have limited our presentation mainly to linear problems. Non-linear matrix equations have a crucial and ever increasing role in many applications: for instance, the popular algebraic Riccati equation (see [168]) has a leading position in control applications and is an important tool in eigenvalue problems; we refer to [43] for a very recent presentation of the rich literature on computational methods. Other fully nonlinear equations include, e.g, equations of the type $\mathbf{X} + A^\top F(\mathbf{X})A = Q$, where F is a properly defined nonlinear function of $\mathbf{X}$; see, e.g., [213] and references therein. Linear matrix equations with special properties arise when dealing with *periodic* dynamical systems. These problems give rise to periodic counterparts of the equations we have analyzed, such as Lyapunov and Sylvester equations. Corresponding Schur forms can be used for their solution, and necessary and sufficient conditions for a periodic discrete-time system to be equivalent to a time-invariant systems are known; for more details on both the theoretical and algorithmic aspects, mainly on small size problems, we refer, e.g., to the contributions of R. Byers, P. Van Dooren, J. Sreedhar, A. Varga on the subject.

10. Acknowledgements. We would like to thank Ulrike Baur, Peter Benner, Paul van Dooren, Lars Grasedyck, Yiding Lin and Daniel B. Szyld, for helpful comments and pointers to many references. We are indebted with Daniel Szyld for providing us with [130]. We are grateful to Daniel Kressner and Froilán M. Dopico for carefully reading and amending a previous version of this manuscript, and for several helpful references. We are also grateful to the librarians of the Math Library at the Dipartimento di Matematica, Università di Bologna, for their professional dedication.

References

[1] M. Abramowitz and I. A. Stegun. *Handbook of Mathematical Functions*. Dover, 1965. [10, 28]

[2] P.-A. Absil, R. Mahony, and R. Sepulchre. *Optimization Algorithms on Matrix Manifolds*. Princeton University Press, Princeton, NJ, 2008. [32]

[3] Mian Ilyas Ahmad, Imad Jaimoukha, and Michalis Frangos. Krylov subspace restart scheme for solving large-scale Sylvester equations. In *2010 American Control Conference*, Marriott Waterfront, Baltimore, MD, USA, June 30-July 02 2010. [18]

[4] A. C. Antoulas. *Approximation of large-scale Dynamical Systems*. Advances in Design and Control. SIAM, Philadelphia, 2005. [4, 5, 15, 18, 32]

[5] A. C. Antoulas, D. C. Sorensen, and Y. Zhou. On the decay rate of Hankel singular values and related issues. *Systems & Control Letters*, 46(5):323–342, August 2002. [21]

[6] Ivo Babuska, Raul Tempone, and Georgios E. Zouraris. Galerkin finite element approximations of stochastic elliptic partial differential equations. *SIAM J. Numer. Anal.*, 42(2):800–825, 2004. [40]

[7] Z. Bai and J. Demmel. Design of a parallel nonsymmetric eigenroutine toolbox, Part I. In R.F. Sincovec et al., editor, *Proc. of the 6th SIAM Conf. on Parallel Processing for Scientific Computing*, 1993. Also as Computer Science Division Tech Report UCB CSD-92-718, U. C. Berkeley, 1992. [23]

[8] Zhaojun Bai and Daniel Skoogh. A projection method for model reduction of bilinear dynamical systems. *Linear Algebra and its Applications*, 415(2-3):406–425, 2006. Special Issue on Order Reduction of Large-Scale Systems. [39]

[9] Zhong-Zhi Bai. On Hermitian and skew-Hermitian splitting iteration methods for continuous Sylvester equations. *Journal of Computational Mathematics*, 29(2):185–198, 2011. [11]

[10] A. D. Barbour and S. Utev. Solving the Stein equation in compound Poisson approximation. *Adv. Appl. Probab.*, 30:449–475, 1998. [6, 32]

[11] J. B. Barlow, M. M. Monahemi, and D. P. O'Leary. Constrained matrix Sylvester equations. *SIAM J. Matrix Anal. Appl.*, 13(1):1–9, 1992. [38]

[12] S. Barnett. Simplification of the Lyapunov matrix equation $A^TPA - P = -Q$. *Automatic Control, IEEE Transactions on*, 19(4):446 – 447, Aug. 1974. [33]

[13] A.Y. Barraud. A numerical algorithm to solve $A^TXA - X = Q$. *IEEE Trans. Automatic Control*, AC-22(5):883–885, October 1977. [33]

[14] J. F. Barrett. Bibliography of A.M. Lyapunov's work. *Int. J. Control*, 55(3):785–790, 1992. [2]

[15] R. H. Bartels and G. W. Stewart. Algorithm 432: Solution of the Matrix Equation $AX + XB = C$. *Comm. of the ACM*, 15(9):820–826, 1972. [2, 10, 11]

[16] Ulrike Baur. *Control-Oriented Model Reduction for Parabolic Systems*. PhD thesis, Institute für Mathematik, TU Berlin, 2008. [30]

[17] Ulrike Baur. Low rank solution of data-sparse Sylvester equations. *Numerical Linear Algebra with Applications*, 15(9):837–851, 2008. [20]

[18] Ulrike Baur and Peter Benner. Factorized solution of Lyapunov equations based on hierarchical matrix arithmetic. *Computing*, 78:211–234, 2006. [23, 30, 31]

[19] Ulrike Baur and Peter Benner. Cross-Gramian based model reduction for data-sparse systems. *Electronic Transactions on Numerical Analysis*, 31:256–270, 2008. [20]

[20] B. Beckermann and L. Reichel. Error estimation and evaluation of matrix functions via the Faber transform. *SIAM J. Numer. Anal.*, 47(5):3849–3883, 2009. [18]

[21] Bernhard Beckermann. An Error Analysis for Rational Galerkin Projection Applied to the Sylvester Equation. *SIAM J. Numer. Anal.*, 49(6):2430–2450, 2011. [18]

[22] P. Benner. Factorized solution of Sylvester equations with applications in control. In B. De Moor, B. Motmans, J. Willems, P. Van Dooren, and V. Blondel, editors, *Proc. Sixteenth Intl. Symp. Math. Theory Networks and Syst.*, MTNS 2004, Leuven (B), July 5-9, 2004. [5, 11]

[23] P. Benner. *Control Theory*, chapter 57. Chapman & Hall/CRC, 2006. Handbook of Linear Algebra. [4]

[24] P. Benner, T. Breiten, and T. Damm. Generalised tangential interpolation for model reduction of discrete-time MIMO bilinear systems. *International Journal of Control*, 84(8):1398–1407, 2011. [39]

[25] P. Benner, P. Ezzatti, D. Kressner, E. S. Quintana-Ortí, and A. Rem'on. A mixed-precision algorithm for the solution of Lyapunov equations on hybrid CPU-GPU platforms. *Parallel Computing*, 37(8):439–450, 2011. [42]

[26] P. Benner, P. Kürschner, and J. Saak. Efficient handling of complex shift parameters in the low-rank Cholesky factor ADI method. *Numerical Algorithms*, 62(2):225–251, 2013. [27, 28]

[27] P. Benner, R.-C. Li, and N. Truhar. On the ADI method for Sylvester equations. *Journal of Computational and Applied Mathematics*, 233(4):1035–1045, 2009. [19]

[28] P. Benner, V. Mehrmann, and D. Sorensen (eds). *Dimension Reduction of Large-Scale Systems*. Lecture Notes in Computational Science and Engineering. Springer-Verlag, Berlin/Heidelberg, 2005. [4, 32]

[29] P. Benner and E. S. Quintana-Ortí. Solving stable generalized Lyapunov equations with the matrix sign function. *Numerical Algorithms*, 20(1):75–100, 1999. [23, 30, 36]

[30] P. Benner and J. Saak. A Galerkin-Newton-ADI method for solving large-scale algebraic Riccati equations. Technical Report SPP1253-090, Deutsche Forschungsgemeinschaft - Priority Program 1253, 2010. [29]

[31] Peter Benner and Tobias Breiten. Low Rank Methods for a Class of Generalized Lyapunov Equations and Related Issues. *Numerische Mathematik*, 123:xx–xx, 2013. to appear. [39]

[32] Peter Benner and Tobias Damm. Lyapunov equations, energy functionals, and model order reduction of bilinear and stochastic systems. *SIAM J. Control Optim.*, 49(2):686–711, 2011. [39]

[33] Peter Benner, Grece El Khoury, and Miloud Sadkane. On the squared Smith method for large-scale Stein equations. Technical Report MPIMD/12-15, Max Planck Institute Magdeburg, September 2012. [34]

[34] Peter Benner, Martin Køhler, and Jens Saak. Sparse-dense Sylvester equations in $\mathcal{H}_2$-model order reduction. Technical Report MPIMD/11-11, Max Planck Institute Magdeburg Preprint, December 2011. Available from `http://www.mpi-magdeburg.mpg.de/preprints/`. [13]

[35] Peter Benner, Daniel Kressner, Vasile Sima, and Andras Varga. The SLICOT Toolboxes - a survey. SLICOT Working Note 2009-1, August 2009. [11, 41]

[36] Peter Benner, Jing-Rebecca Li, and Thilo Penzl. Numerical solution of large-scale Lyapunov equations, Riccati equations, and linear-quadratic optimal control problems. *Num. Lin. Alg. with Appl.*, 15:1–23, 2008. [27]

[37] Peter Benner, Enrique S. Quintana-Ortì, and Gregorio Quintana-Ortì. Solving stable Stein equations on distributed memory computers. In P. Amestoy, P. Berger, M. Daydè, I. Duff, V. Frayssè, L. Giraud, and D. Ruiz, editors, *EuroPar'99 Parallel Processing*, number 1685 in Lecture Notes in Computer Science, pages 1120–1123. Springer-Verlag, 1999. [42]

[38] Peter Benner, Enrique S. Quintana-Ortì, and Gregorio Quintana-Ortì. Numerical solution of discrete stable linear matrix equations on multicomputers. *Parallel Algorithms and Applications*, 17(1):127–146, 2002. [42]

[39] Peter Benner, Enrique S. Quintana-Ortì, and Gregorio Quintana-Ortì. Solving stable Sylvester equations via rational iterative schemes. *Journal of Scientific Computing*, 28(1):51–83, 2006. [11]

[40] A. Betser, N. Cohen, and E. Zeheb. On solving the Lyapunov and Stein equations for a companion matrix. *Systems & Control Letters*, 25(3):211–218, June 1995. [22, 33]

[41] J.H. Bevis, F.J. Hall, and R.E. Hartwig. The matrix equation $AX - XB = C$ and its special cases. *SIAM Journal on Matrix Analysis and Applications*, 9(3):348–359, 1988. [40]

[42] W. G. Bickley and J. McNamee. Matrix and other direct methods for the solution of linear difference equation. *Philos. Trans. Roy. Soc. London Ser. A*, 252:69–131, 1960. [12]

[43] D.A. Bini, B. Iannazzo, and B. Meini. *Numerical Solution of Algebraic Riccati Equations*. SIAM, 2012. [5, 6, 42]

[44] Dario A. Bini, Beatrice Meini, and Federico Poloni. Fast solution of a certain Riccati equation through Cauchy-like matrices. *Electron. Trans. Numer. Anal.*, 33:84–104, 2009. [12]

[45] Christian H. Bischof, Biswa N. Datta, and Avijit Purkayastha. A Parallel Algorithm for the Multi-input Sylvester-Observer Equation. In *American Control Conference*, pages 152 – 156, 1992. [42]

[46] Matthias Bollhöfer and André K. Eppler. A structure preserving FGMRES method for solving large Lyapunov equations. In M. Günther, A. Bartel, M. Brunk, S. Schoeps, and M. Striebel, editors, *Progress in Industrial Mathematics at ECMI 2010, Mathematics in Industry,*, volume 17, pages 131–136, 2012. [30]

[47] I. Borno. Parallel computation of the solutions of coupled algebraic Lyapunov equations. *Automatica*, 31(9):1345–1347, September 1995. [37, 42]

[48] A. Bouhamidi and K. Jbilou. A note on the numerical approximate solutions for generalized Sylvester matrix equations with applications. *Applied Math. Comp.*, 206:687–694, 2008. [35]

[49] H. W. Braden. The equations $A^T X \pm X^T A = B$. *SIAM J. Matrix Anal. Appl.*, 20(2):295–302, 1998. [40]

[50] Sanjoy Brahma and Biswa Datta. An optimization approach for minimum norm and robust partial quadratic eigenvalue assignment problems for vibrating structures. *Journal of Sound and Vibration*, 324(3-5):471–489, 2009. [5]

[51] R. Byers. A LINPACK-style condition estimator for the equation $AX + XB^T = C$. *IEEE Trans. Automat. Control*, AC-29:926–928, 1984. [9]

[52] R. Byers. Solving the algebraic Riccati equation with the matrix sign function. *Linear Algebra Appl.*, 85:267–279, 1987. [23]

[53] R. Byers and D. Kressner. Structured condition numbers for invariant subspaces. *SIAM J. Matrix Anal. Appl.*, 28(2):326–347, 2006. [40]

[54] D. Calvetti, B. Lewis, and L. Reichel. On the solution of large Sylvester-observer equations. *Numer. Linear Algebra Appl.*, 8:435–451, 2001. [38]

[55] D. Calvetti and L. Reichel. Application of ADI iterative methods to the restoration of noisy images. *SIAM J. Matrix Anal. Appl.*, 17(1):165–186, January 1996. [5]

[56] J. Carvalho and B. A. Datta. A block algorithm for the Sylvester-observer equation arising in state-estimation. In *Proceedings IEEE International Conference on Decision and Control*, pages 3898–3903, Orlando, 2001. [38]

[57] J. Carvalho, K. Datta, and Yoopyo Hong. A new block algorithm for full-rank solution of the Sylvester-observer equation. *Automatic Control, IEEE Transactions on*, 48(12):2223–2228, Dec. 2003. [38]

[58] Eugenio B. Castelan and Vilemar Gomes da Silva. On the solution of a Sylvester equation appearing in descriptor systems control theory. *Systems & Control Letters*, 54:109–117, February 2005. [38]

[59] Chong Chen and Dan Schonfeld. Pose estimation from multiple cameras based on Sylvester's equation. *Computer Vision and Image Understanding*, 114(6):652–666, June 2010. [5, 6]

[60] Chong Chen, Dan Schonfeld, Junlan Yang, and Magdi Mohamed. Pose estimation from video sequences based on Sylvester's equation. In *SPIE Proceedings of Electronic Imaging, Conference on Visual Communications and Image Processing*, San Jose, CA, 2007. SPIE'07. [5]

[61] Benchmark Collection. Oberwolfach model reduction benchmark collection, 2003. `http://www.imtek.de/simulation/benchmark`. [41]

[62] Marisa Condon and Rossen Ivanov. Nonlinear systems - algebraic Gramians and model reduction. *Compel*, 24(1):202–219, 2005. [5, 39]

[63] M. J. Corless and A. E. Frazho. *Linear systems and control - An operator perspective*. Pure and Applied Mathematics. Marcel Dekker, New York - Basel, 2003. [4]

[64] O. L. V. Costa and C. S. Kubrusly. Lyapunov equation for infinite-dimensional discrete bilinear systems. *Systems & Control Letters*, 17(1):71–77, July 1991. [39]

[65] T. Damm. Direct methods and ADI-preconditioned Krylov subspace methods for generalized Lyapunov equations. *Num. Lin. Alg. with Appl.*, 15:853–871, 2008. Special issue on Matrix equations. [39]

[66] M. Darouach. Solution to Sylvester equation associated to linear descriptor systems. *Systems & Control Letters*, 55(10):835–838, October 2006. [38]

[67] B. N. Datta and Y. Saad. Arnoldi methods for large Sylvester-like observer matrix equations, and an associated algorithm for partial spectrum assignment. *Lin. Alg. Appl.*, 154–156:225–244, 1991. [38]

[68] Biswa Nath Datta. *Numerical methods for linear control systems: design and analysis*. Elsevier, Academic Press, 2004. [41]

[69] Biswa Nath Datta and Chandanie Hetti. Generalized Arnoldi methods for the Sylvester-observer equation and the multi-input pole placement problem. In *Proceedings of the 36th Conference on Decision & Control*, pages 4379–4383, San Diego, California USA, December 1997. [38]

[70] B.N. Datta. Linear and numerical linear algebra in control theory: Some research problems. *Lin. Alg. Appl.*, 197–198:755–790, 1994. [4, 38]

[71] B.N. Datta, M. Heyouni, and K. Jbilou. The global Arnoldi process for solving the Sylvester-observer equation. *Computational and Applied Mathematics*, 29(3):527–544, 2010. [38]

[72] E. de Souza and S. P. Bhattacharyya. Controllability, Observability and the Solution of $AX - XB = C$. *Lin. Alg. Appl.*, 39:167–188, 1981. [7, 8]

[73] J. Demmel. Three methods for refining estimates of invariant subspaces. *Computing*, 38:43–57, 1987. [5]

[74] Feng Ding and Tongwen Chen. Iterative least-squares solutions of coupled Sylvester matrix equations. *Systems & Control Letters*, 54(2):95–107, February 2005. [37]

[75] P. Van Dooren. Reduced order observers: a new algorithm and proof. *System and Control Letters*, 4:243–251, 1984. [38]

[76] V. Druskin and L. Knizhnerman. Extended Krylov subspaces: approximation of the matrix square root and related functions. *SIAM J. Matrix Anal. Appl.*, 19(3):755–771, 1998. [17]

[77] V. Druskin, L. Knizhnerman, and M. Zaslavsky. Solution of large scale evolutionary problems using rational Krylov subspace optimized shifts. *SIAM J. Sci. Comput.*, 31(5):3760–3780, 2009. [18]

[78] V. Druskin and V. Simoncini. Adaptive rational Krylov subspaces for large-scale dynamical systems. *Systems and Control Letters*, 60:546–560, 2011. [25, 26, 28]

[79] V. Druskin, V. Simoncini, and M. Zaslavsky. Adaptive tangential interpolation in rational Krylov subspaces for MIMO model reduction data. Technical report, Dipartimento di Matematica, Università di Bologna, November 2012. [26]

[80] Vladimir Druskin, Leonid Knizhnerman, and V. Simoncini. Analysis of the rational Krylov subspace and ADI methods for solving the Lyapunov equation. *SIAM J. Numer. Anal.*, 49:1875–1898, 2011. [29]

[81] Guangren Duan. A note on combined generalized Sylvester matrix equations. *Journal of Control Theory and Applications*, 4:397–400, 2004. [37]

[82] M. Eiermann and O. Ernst. A restarted Krylov subspace method for the evaluation of matrix functions. Technical report, Institut für Numerische Mathematik und Optimierung, Technische Universität Bergakademie Freiberg, 2005. SIAM J. Numerical Analysis. To appear. [18]

[83] Nancy S. Ellner and Eugene L. Wachspress. New ADI model problem applications. In *Proceedings of 1986 ACM Fall joint computer conference, Dallas, Texas, United States*, pages 528–534. IEEE Computer Society Press, Los Alamitos, CA, 1986. [6, 10, 15, 18, 19]

[84] Nancy S. Ellner and Eugene L. Wachspress. Alternating Direction Implicit iteration for systems with complex spectra. *SIAM J. Numer. Anal.*, 23(3):859–870, 1991. [10, 28]

[85] E. Emre and L. M. Silverman. The equation $xr + qy = \phi$: A characterization of solutions. *SIAM J. Control Optim.*, 19(1):3338, 1981. [36]

[86] M. Epton. Methods for the solution of $ADXD - BXC = E$ and its application in the numerical solution of implicit ordinary differential equations. *BIT*, 20:341–345, 1980. [6]

[87] O.G. Ernst, C.E. Powell, D.J. Silvester, and E. Ullmann. Efficient solvers for a linear stochastic Galerkin mixed formulation of diffusion problems with random data. *SIAM J. Sci. Comput.*, 31:1424–1447, 2009. [37]

[88] K. V. Fernando and H. Nicholson. On the structure of balanced and other principal representations of SISO systems. *Automatic Control, IEEE Transactions on*, 28(2):228–231, 1983. [5]

[89] K. V. Fernando and H. Nicholson. On a fundamental property of the cross-Gramian matrix. *IEEE Trans. Circ. Syst.*, CAS-31(5):504–505, 1984. [5]

[90] A. Frommer and V. Simoncini. Matrix functions. In Wil H. A. Schilders, Henk A. van der Vorst, and Joost Rommes, editors, *Model Order Reduction: Theory, Research Aspects and Applications*, Mathematics in Industry. Springer, Heidelberg, 2008. [23, 24]

[91] Z. Gajić and M. Qureshi. *Lyapunov Matrix Equation in System Stability and Control*. Math. in Science and Engineering. Academic Press, San Diego, CA, 1995. [4, 5, 6, 8, 21, 22]

[92] K. Gallivan, A. Vandendorpe, and P. Van Dooren. Sylvester equations and projection-based model reduction. *J. Comput. Appl. Math.*, 162(1):213–229, 2004. [5]

[93] Stephan Ramon Garcia and Amy L. Shoemaker. On the matrix equation $AX+XA^T = 0$. *Lin. Alg. Appl.*, 438:2740–2746, 2013. In Press. [41]

[94] Judith D. Gardiner, Alan J. Laub, James J. Amato, and Cleve B. Moler. Solution of the Sylvester matrix equation $AXB^{\mathrm{T}} + CXD^{\mathrm{T}} = E$. *ACM Transactions on Mathematical Software*, 18(2):223–231, June 1992. [35]

[95] Judith D. Gardiner, Matthew R. Wette, Alan J. Laub, James J. Amato, and Cleve B. Moler. Algorithm 705: A FORTRAN-77 software package for solving the Sylvester matrix equation $AXB^{\mathrm{T}} + CXD^{\mathrm{T}} = E$. *ACM Transactions on Mathematical Software*, 18(2):232–238, June 1992. [35]

[96] Ali R. Ghavimi and Alan J. Laub. Numerical methods for nearly singular constrained matrix Sylvester equations. *SIAM J. Matrix Anal. Appl.*, 17:212–221, 1996. [38]

[97] I. Gohberg, T. Kailath, and V. Olshevsky. Fast Guassian elimination with partial pivoting for matrices with displacement structure. *Math. Comp.*, 64(212):1557–1576, 1995. [12]

[98] I. Gohberg, P. Lancaster, and L. Rodman. *Invariant Subspaces of Matrices with Applications*. J. Wiley & Sons, New York, 1986. [7]

[99] G. Golub and C. F. Van Loan. *Matrix Computations*. The Johns Hopkins University Press, Baltimore, 3rd edition, 1996. [9, 11]

[100] G. H. Golub, S. Nash, and C. Van Loan. A Hessenberg-Schur method for the problem $AX + XB = C$. *IEEE Trans. Autom. Contr.*, AC-24(6):909–913, 1979. [11]

[101] Robert Granat. Parallel algorithms and condition estimators for standard and generalized triangular sylvester-type matrix equations. In *PARA06 - State of the Art in Scientific and Parallel Computing. Vol. 4699. Lecture Notes in Computer Science*. Springer, 2007. [42]

[102] L. Grasedyck. Existence of a low rank or $\mathcal{H}$-matrix approximant to the solution of a Sylvester equation. *Numerical Linear Algebra with Appl.*, 11(4):371–389, 2004. [20]

[103] L. Grasedyck. Nonlinear multigrid for the solution of large-scale Riccati equations in low-rank and $\mathcal{H}$-matrix format. *Numer. Linear Algebra Appl.*, 15:779–807, 2008. [31]

[104] L. Grasedyck and W. Hackbusch. A multigrid method to solve large scale Sylvester equations. *SIAM J. Matrix Anal. Appl.*, 29:870–894, 2007. [20]

[105] L. Grasedyck, W. Hackbusch, and B. Khoromskij. Solution of large scale algebraic matrix Riccati equations by use of hierarchical matrices. *Computing*, 70(2):121–165, 2003. [31]

[106] Lars Grasedyck. *Theorie und Anwendungen Hierarchischer Matrizen*. PhD thesis, Mathematisch-Naturwissenschaftliche Fakultät, Christian-Albrechts-Universität zu Kiel, 2001. [20]

[107] W. S. Gray and J. Mesko. Energy functions and algebraic Gramians for bilinear systems. In *Preprints of the 4th IFAC Nonlinear Control Systems Design Symposium*, Enschede, The Netherlands, 1998. [39]

[108] V. Grimm and M. Hochbruck. Rational approximation to trigonometric operators. *BIT*, 48:215–229, 2008. [18]

[109] E. Grimme. *Krylov projection methods for model reduction*. PhD thesis, The University of Illinois at Urbana-Champaign, 1997. [18]

[110] T. Gudmundsson and A. J. Laub. Approximate solution of large sparse Lyapunov equations. *IEEE Trans. Automatic Control*, 39(5):1110–1114, 1994. [31]

[111] A. El Guennouni, K. Jbilou, and A.J. Riquet. Block Krylov subspace methods for solving large Sylvester equations. *Numerical Algorithms*, 29:75–96, 2002. [13, 15]

[112] S. Gugercin, A. C. Antoulas, and C. Beattie. $\mathcal{H}_2$ model reduction for large-scale linear dynamical systems. *SIAM J. Matrix Anal. Appl.*, 30:609–638, 2008. [29]

[113] S. Gugercin, D. C. Sorensen, and A. C. Antoulas. A modified low-rank Smith method for large-scale Lyapunov equations. *Numerical Algorithms*, 32:27–55, 2003. [27]

[114] M. H. Gutknecht. Krylov subspace algorithms for systems with multiple right hand sides: a collection of references, 2004. available at `http://www.sam.math.ethz.ch/~mhg/talks/abstr/abstr.html`. [3]

[115] Stefan Güttel. *Rational Krylov Methods for Operator Functions.* PhD thesis, TU Bergakademie Freiberg, Germany, 2010. [18]

[116] N. Hale, N. Higham, and L. N. Trefethen. Computing A^α, $\log(A)$ and Related Matrix Functions by Contour Integrals. *SIAM J. Numer. Anal.*, 46(5):2505–2523, 2008. [18]

[117] S. J. Hammarling. Numerical solution of the stable, non-negative definite Lyapunov equation. *IMA J. Numer. Anal.*, 2:303–323, 1982. [22, 36]

[118] Sven Hammarling. Numerical solution of the discrete-time, convergent, non-negative definite Lyapunov equation. *Systems & Control Letters*, 17(2):137–139, August 1991. [33]

[119] C. Hartmann, A. Zueva, and B. Schäfer-Bung. Balanced model reduction of bilinear systems with applications to positive systems. Technical report, Institut für Mathematik, Freie Universität, Berlin, 2010. [39]

[120] He Kong and Bin Zhou and Mao-Rui Zhang. A Stein equation approach for solutions to the Diophantine equations. In *Control and Decision Conference (CCDC), 2010 Chinese*, pages 3024 – 3028, 2010. [32, 41]

[121] J. Z. Hearon. Nonsingular Solutions of $TA - BT = C$. *Lin. Alg. Appl.*, 16:57–63, 1977. [8]

[122] G. Hewer and C. Kenney. The sensitivity of the stable Lyapunov equation. *SIAM J. Control and Optimization*, 26:321344, 1988. [21]

[123] M. Heyouni and K. Jbilou. Matrix Krylov subspace methods for large scale model reduction problems. *Applied Mathematics and Computation*, 181:1215–1228, October 2006. [26]

[124] N. J. Higham. Perturbation theory and backward error for $AX - XB = C$. *BIT*, 33:124–136, 1993. [9, 10, 21]

[125] Nicholas J. Higham. *Matrix Functions – Theory and Applications.* SIAM, Philadelphia, USA, 2008. [18]

[126] M. Hochbruck and C. Lubich. On Krylov subspace approximations to the matrix exponential operator. *SIAM J. Numer. Anal.*, 34(5):1911–1925, 1997. [18]

[127] M. Hochbruck and G. Starke. Preconditioned Krylov subspace methods for Lyapunov matrix equations. *SIAM Matrix Anal. and Appl.*, 16(1):156–171, 1995. [30]

[128] A. Hodel, K. Poolla, and B. Tenison. Numerical solution of the Lyapunov equation by approximate power iteration. *Lin. Alg. Appl.*, 236:205–230, 1996. [31]

[129] A. Scottedward Hodel and Pradeep Misra. Least-squares approximate solution of overdetermined Sylvester equations. *SIAM J. Matrix Anal. Appl.*, 18(2):279–290, April 1997. [16]

[130] Alan Scottedward Hodel. *Numerical methods for the solution of large and very large, sparse Lyapunov equations.* PhD thesis, Engineering, Electronics and Electrical, University of Illinois at Urbana-Champaign, 1989. [24, 42]

[131] A.Scottedward Hodel and K. Poolla. Numerical solution of very large, sparse Lyapunov equations through approximate power iteration. In *Proceedings of the 29th Conference on Decision and Control, IEEE, WA12*, pages 291–296, December 1990. [31]

[132] R. A. Horn and C. R. Johnson. *Topics in Matrix Analysis*. Cambridge University Press, Cambridge, 1991. [7, 21, 37]

[133] D. Y. Hu and L. Reichel. Krylov Subspace Methods for the Sylvester Equation. *Linear Algebra and Appl.*, 172:283–313, July 1992. [16, 17, 18, 23, 26]

[134] Guang-Da Hu and Qiao Zhu. Bounds of modulus of eigenvalues based on Stein equation. *Kybernetika*, 46(4):655–664, 2010. [32]

[135] J. L. Hueso, G. Martínez, and V. Hernández. A systolic algorithm for the triangular Stein equation. *Journal of VLSI signal processing systems for signal, image and video technology*, 5(1):49–55, 1993. [33]

[136] Kh. D. Ikramov. Conditions for unique solvability of the matrix equation $AX+X^TB = C$. *Doklady Mathematics*, 81(1):63–65, 2010. [40]

[137] C. Jagels and L. Reichel. The extended Krylov subspace method and orthogonal Laurent polynomials. *Lin. Alg. Appl.*, 431:441–458, 2009. [18]

[138] I. M. Jaimoukha and E. M. Kasenally. Oblique projection methods for large scale model reduction. *SIAM J. Matrix Anal. Appl.*, 16(2):602–627, 1995. [26]

[139] I. M. Jaimoukha and E. M. Kasenally. Krylov subspace methods for solving large Lyapunov equations. *SIAM J. Numer. Anal.*, 31(1):227–251, Feb. 1994. [24, 25, 26]

[140] Imad M. Jaimoukha and Ebrahim M. Kasenally. Implicitly Restarted Krylov Subspace Methods for Stable Partial Realizations. *SIAM J. Matrix Anal. Appl.*, 18(3):633–652, 1997. [26]

[141] Antony Jameson. Solution of the equation $AX + XB = C$ by inversion of an $M \times M$ or $N \times N$ matrix. *SIAM J. Appl. Math.*, 16(5):1020–1023, 1968. [7, 8]

[142] K. Jbilou. Low rank approximate solutions to large Sylvester matrix equations. *Applied Mathematics and Computation*, 177:365–376, 2006. [17, 33]

[143] K. Jbilou. ADI preconditioned Krylov methods for large Lyapunov matrix equations. *Lin. Alg. Appl.*, 432:2473–2485, May 2010. [29]

[144] K. Jbilou, A. Messaoudi, and H. Sadok. Global FOM and GMRES algorithms for matrix equations. *Applied Numerical Mathematics*, 31:49–63, 1999. [17]

[145] K. Jbilou and A.J. Riquet. Projection methods for large Lyapunov matrix equations. *Linear Algebra and its Applications*, 415(2-3):344–358, 2006. [26]

[146] I. Jonsson and B. Kågström. Recursive blocked algorithms for solving triangular systems Part I: One-sided and coupled Sylvester-type matrix equations. *ACM Trans. Math. Softw.*, 28(4):392–415, 2002. [11, 42]

[147] I. Jonsson and B. Kågstrøm. Recursive blocked algorithms for solving triangular systems Part II: Two-sided and generalized Sylvester and Lyapunov matrix equations. *ACM Trans. Math. Softw.*, 28(4):416–435, 2002. [11, 42]

[148] K. Jbilou and A. Messaoudi. A Computational Method for Symmetric Stein Matrix Equations. In *Numerical Linear Algebra in Signals, Systems and Control*, volume 80 of *Lecture Notes in Electrical Engineering*, pages 295–311. Springer, 2011. [33]

[149] A. Kaabi, A. Kerayechian, and F. Toutounian. A new version of successive approximations method for solving Sylvester matrix equations. *Applied Mathematics and Computation*, 186:638–645, 2007. [15]

[150] B. Kågström and P. Poromaa. LAPACK-Style algorithms and software for solving the generalized Sylvester equation and estimating the separation between regular matrix pairs. *ACM Transactions on Mathematical Software*, 22(1):78–103, 1996. Also as LAPACK Working Note 75. [37]

[151] Bo Kågström and Lars Westin. Generalized Schur methods with Condition Estimators for Solving the Generalized Sylvester equation. *Automatic Control, IEEE Transactions on*, 34:745–751, 1989. [37]

[152] Thomas Kailath. *Linear systems*. Englewood Cliffs, N.J.; Prentice-Hall, 1980. [33]

[153] A. Klein and P. Spreij. On Fisher's information matrix of an ARMAX process and Sylvester's resultant matrices. *Linear Algebra Appl.*, 237/238:579–590, 1996. [6]

[154] A. Klein and P.J.C. Spreij. On Stein's equation, Vandermonde matrices and Fisher's information matrix of time series processes. Part I: The autoregressive moving average process. *Lin. Alg. Appl.*, 329(1-3):9–47, 2001. [6, 32]

[155] André Klein and Peter Spreij. On the solution of Stein's equation and Fisher's information matrix of an ARMAX process. *Linear Algebra and its Applications*, 396:1–34, 2005. [6, 32]

[156] L. Knizhnerman and V. Simoncini. A new investigation of the extended Krylov subspace method for matrix function evaluations. *Numerical Linear Algebra with Applications*, 17(4):615–638, 2010. [18]

[157] L. Knizhnerman and V. Simoncini. Convergence analysis of the Extended Krylov Subspace Method for the Lyapunov equation. *Numerische Mathematik*, 118(3):567–586, 2011. [25]

[158] N. Komaroff. Simultaneous eigenvalue lower bounds for the Lyapunov matrix equation. *IEEE Trans. on Autom. Control*, 33(1):126–128, 1988. [21]

[159] M. Konstantinov, D. Gu, V. Mehrmann, and P. Petkov. *Perturbation theory for matrix equations*. Studies in Computational Mathematics 9. Elsevier, 2003. [3, 9]

[160] Krener, A. J. Reduced Order Models for Nonlinear Control Systems. In A. Astolfi and L. Marconi, editors, *Analysis and Design of Nonlinear Control Systems*. Springer Verlag, 2007. In Honor of Alberto Isidori. [5]

[161] D. Kressner. Block variants of Hammarling's method for solving Lyapunov equations. *ACM Trans. Math. Software*, 34(1):1–15, 2008. [22, 33]

[162] Daniel Kressner. Memory-efficient Krylov subspace techniques for solving large-scale Lyapunov equations. In *IEEE International Symposium on Computer-Aided Control Systems*, pages 613–618, San Antonio, 2008. [18]

[163] Daniel Kressner, Volker Mehrmann, and Thilo Penzl. CTLEX - a collection of benchmark examples for continuous-time Lyapunov equations. SLICOT Working Note 1999-6, May 1999. [41]

[164] Daniel Kressner, Christian Schröder, and David S. Watkins. Implicit QR algorithms for palindromic and even eigenvalue problems. *Numer. Algor.*, 51:209–238, 2009. [40]

[165] Daniel Kressner and Christine Tobler. Low-rank tensor Krylov subspace methods for parametrized linear systems. *SIAM. J. Matrix Anal. & Appl.*, 32(4):1288–1316, 2011. [39]

[166] Peter Kunkel and Volker Ludwig Mehrmann. *Differential-Algebraic Equations: Analysis and Numerical Solution*. European Mathematical Society, 2006. [3, 35]

[167] P. Lancaster, L. Lerer, and M. Tismenetsky. Factored forms for solutions of $AX - XB = C$ and $X - AXB = C$ in companion matrices. *Lin. Alg. Appl.*, 62:19–49, 1984. [8, 33]

[168] P. Lancaster and L. Rodman. *Algebraic Riccati equations*. Oxford Univ. Press, 1995. [32, 42]

[169] P. Lancaster and M. Tismenetsky. *The Theory of Matrices*. Academic Press, New York, second edition, 1985. [7, 21, 32]

[170] Peter Lancaster. Explicit solutions of linear matrix equations. *SIAM Review*, 12(4):544–566, 1970. [3, 7, 8]

[171] V. Larin and F. Aliev. Construction of square root factor for solution of a Lyapunov matrix equation. *Systems Control Lett.*, 20:109–112, 1993. [23]

[172] A. Laub. Numerical linear algebra aspects of control design computations. *IEEE Trans. Automat. Control*, AC-30:97–108, 1985. [5]

[173] R. B. Lehoucq, D. C. Sorensen, and C. Yang. *ARPACK Users Guide: Solution of large scale Eigenvalue Problems by Implicitly Restarted Arnoldi Methods*. SIAM, Philadelphia, 1998. [31]

[174] L. Lerer and A.C.M. Ran. A new inertia theorem for Stein equations, inertia of invertible Hermitian block Toeplitz matrices and matrix orthogonal polynomials. *Integral Equations and Operator Theory*, 47:339–360, 2003. [32]

[175] J.-R. Li and J. White. Low-Rank solutions of Lyapunov equations. *SIAM J. Matrix Anal. Appl.*, 24(1):260–280, 2002. [27, 28]

[176] Jing-Rebecca Li. *Model Reduction of Large Linear Systems via Low Rank System Gramians*. PhD thesis, MIT, 2000. [24]

[177] Tiexiang Li, Chang-Yi Weng, Eric King wah Chu, and Wen-Wei Lin. Solving Large-Scale Stein and Lyapunov Equations by Doubling. Technical Report 2012-05-005, Math.Division, NCTS, Taiwan, 2012. [34]

[178] Tiexiang Li, Peter Chang-Yi Weng, Eric King wah Chu, and Wen-Wei Lin. Large-scale Stein and Lyapunov equations, Smith method, and applications. *Numer. Algor.*, 123:xx–xx, 1013. To appear. [34]

[179] Joerg Liesen and Zdenek Strakos. *Krylov Subspace Methods. Principles and analysis*. Oxford University Press, 2012. [22]

[180] Matthew M. Lin and Chun-Yueh Chiang. On the T-Stein matrix equation. Technical Report 2012-10-002, National Center for Theoretical Sciences, National Tsing Hua University, Taiwan, 2012. [34]

[181] Y. Lin and V. Simoncini. Minimal residual methods for large scale Lyapunov equations. Technical report, Dipartimento di Matematica, Università di Bologna, 2012. [16, 24, 26]

[182] Yiqin Lin, Liang Bao, Qinghua Wu, and Liping Zhou. A Schur Method for Solving Projected Continuous-Time Sylvester Equations. *International Journal of Computational and Mathematical Sciences*, 4(3), 2010. [39]

[183] Yiqin Lin and Yimin Wei. Condition numbers of the generalized Sylvester equation. *IEEE Transactions on Automatic Control*, 52:2380 – 2385, Dec. 2007. [37]

[184] An Lu and Eugene L. Wachspress. Solution of Lyapunov equations by Alternating Direction Implicit iteration. *Computers Math. Applic.*, 21(9):43–58, 1991. [26]

[185] D. Luenberger. Observers for multivariable systems. *IEEE Trans. Automat. Control*, 11:190–197, 1966. [38]

[186] A. M. Lyapunov. The general problem of the stability of motion. *International Journal of Control*, 55(3):531–534, 1992. Centennial edition of the original article by A. M. Lyapunov published in 1892. [2]

[187] M. Mariton. *Jump linear systems in automatic control*. Marcel Dekker, New York - Basel, 1990. [39]

[188] The MathWorks, Inc. *MATLAB 7*, September 2004. [41]

[189] Karl Meerbergen and Alastair Spence. Inverse iteration for purely imaginary eigenvalues with application to the detection of Hopf bifurcations in large scale problems. *SIAM Journal on Matrix Analysis and Applications*, 31(4):1982–1999, May 2010. [5, 36]

[190] Carl Christian Kjelgaard Mikkelsen. *Numerical methods for large Lyapunov equations*. PhD thesis, Purdue University, West Lafayette, Indiana, May 2009. [24, 30]

[191] C. B. Moler and G. W. Stewart. An algorithm for generalized eigenvalue problems. *SIAM J. Numer. Anal.*, 10:241–256, 1973. [36]

[192] Moghen M. Monahemi, Jewel B. Barlow, and Dianne P. O'Leary. The design of reduced order observers with precise loop transfer recovery. *AIAA Journal of Guidance, Control, and Dynamics*, 15(6):1320–1326, 1992. [38]

[193] B. C. Moore. Principal component analysis in linear systems: Controllability, observability, and model reduction. *Automatic Control, IEEE Transactions on*, AC-26:17–32, Feb. 1981. [4]

[194] I. Moret and P. Novati. RD-rational approximations of the matrix exponential. *BIT, Numerical Mathematics*, 44(3):595–615, 2004. [18]

[195] N. M. Nachtigal, L. Reichel, and L. N. Trefethen. A hybrid GMRES algorithm for nonsymmetric matrix iterations. *SIAM J. Matrix Anal. Appl.*, 13(3):796–825, July 1992. [29]

[196] Hung Dinh Nong. *Numerical Solutions of Matrix Equations Arising in Model Reduction of Large-Scale Linear-Time-Invariant Systems*. PhD thesis, Electrical and Computer Engineering, Rice University, January 2010. [24, 31]

[197] Ryan Nong and Danny C. Sorensen. A parameter free ADI-like method for the numerical solution of large scale Lyapunov equations. Technical report, CAM - Rice University, May 2009. [31]

[198] R. Patel, A. J. Laub, and P. Van Dooren. *Numerical Linear Algebra Techniques for Systems and Control*. IEEE Press, Piscataway NJ, 1993. [5]

[199] D. W. Peaceman and Jr. H. H. Rachford. The numerical solution of parabolic and elliptic equations. *Journal of SIAM*, 3:28–41, 1955. [18]

[200] Ralf Peeters and Paolo Rapisarda. A two-variable approach to solve the polynomial Lyapunov equation. *Systems & Control Letters*, 42(2):117–126, February 2001. [40]

[201] T. Penzl. Numerical solution of generalized Lyapunov equations. *Advances in Computational Mathematics*, 8:33–48, 1998. [35, 36]

[202] T. Penzl. *Numerische Lösung großer Lyapunov-Gleichungen*. Logos-Verlag, Berlin, 1998. Diss. Fakultät für Mathematik, TU Chemnitz. [24]

[203] T. Penzl. A cyclic low-rank Smith method for large sparse Lyapunov equations. *SIAM J. Sci. Comput.*, 21(4):1401–1418, 2000. [27, 28, 34]

[204] T. Penzl. Eigenvalue decay bounds for solutions of Lyapunov equations: the symmetric case. *Systems and Control Letters*, 40(2):139–144, 2000. [10, 21, 22]

[205] T. Penzl. Lyapack users guide. Technical Report SFB393/00-33, TU Chemnitz, 09107 Chemnitz, D, 2000. Available from `http://www.tu-chemnitz.de/sfb393/sfb00pr.html`. [27, 28, 41]

[206] V. Popov. Hyperstability and optimality of automatic systems with several control functions. *Rev. Roum. Sci. Tech.*, 9:929–690, 1964. [33]

[207] Peter Poromaa. Parallel algorithms for triangular Sylvester equations: Design, scheduling and scalability issues. In *Applied Parallel Computing Large Scale Scientific and Industrial Problems*, volume 1541, pages 438–446. Springer, 1998. [41]

[208] H.M. Power. A note on the matrix equation $A'LA - L = -K$. *Automatic Control, IEEE Transactions on*, 14(4):411–412, August 1969. [33]

[209] Hua-Sheng Zhang Qing-Wen Wang and Guang-Jing Song. A new solvable condition for a pair of generalized Sylvester equations. *Electronic Journal of Linear Algebra*, 18:289–301, June 2009. [37]

[210] Daizhan Cheng Qingxi Hu. The polynomial solution to the Sylvester matrix equation. *Applied Mathematics Letters*, 19:859–864, 2006. [12]

[211] E. S. Quintana-Ortí and R. van de Geijn. Specialized parallel algorithms for solving Lyapunov and Stein equations. *J. of Parallel and Distributed Computing*, 61(10):1489–1504, 2001. [42]

[212] E. S. Quintana-Ortí and R. van de Geijn. Formal derivation of algorithms: the triangular Sylvester equation. *ACM Trans. on Mathematical Software*, 29:218–243, 2003. [41]

[213] A. C. M. Ran, M. C. B. Reurings, and L. Rodman. A perturbation analysis for nonlinear selfadjoint operator equations. *SIAM J. Matrix Anal. Appl.*, 28:89–104, 2006. [42]

[214] Hongwu Ren and Richard Dekany. Fast wave-front reconstruction by solving the Sylvester equation with the alternating direction implicit method. *Optics Express*, 12:3279–3296, 2004. [5]

[215] Mickaël Robbé and Miloud Sadkane. A convergence analysis of GMRES and FOM methods for Sylvester equations. *Numerical Algorithms*, 30:71–89, 2002. [14, 15]

[216] J. D. Roberts. Linear model reduction and solution of the algebraic Riccati equation by use of the sign function. *Int. J. Control*, 32(32):677–687, 1980. [23]

[217] William E. Roth. The equations $AX - YB = C$ and $AX - XB = C$ in matrices. *Proceedings of the American Mathematical Society*, 3(3):392–396, Jun. 1952. [7, 36, 37]

[218] A. Ruhe. Rational Krylov sequence methods for eigenvalue computation. *Lin. Alg. Appl.*, 58:391–405, 1984. [18]

[219] Y. Saad. Numerical solution of large Lyapunov equations. In M. A. Kaashoek, J. H. van Schuppen, and A. C. Ran, editors, *Signal Processing, Scattering, Operator Theory, and Numerical Methods. Proceedings of the international symposium MTNS-89, vol III*, pages 503–511, Boston, 1990. Birkhauser. [16, 24]

[220] Y. Saad. *Iterative methods for sparse linear systems*. SIAM, Society for Industrial and Applied Mathematics, 2nd edition, 2003. [3, 13, 15, 18]

[221] John Sabino. *Solution of Large-Scale Lyapunov Equations via the Block Modified Smith Method*. PhD thesis, Rice University, 2006. [10, 22, 24]

[222] Miloud Sadkane. A low-rank Krylov squared Smith method for large-scale discrete-time Lyapunov equations. *Linear Algebra and its Applications*, 436(8):2807–2827, 2012. [34]

[223] B. Schäfer-Bung, C. Hartmann, B. Schmidt, and C. Schütte. Dimension reduction by balanced truncation: Application to light-induced control of open quantum systems. *The Journal of Chemical Physics*, 135:014112, 2011. (13 pages). [39]

[224] W. H. A. Schilders, H. A. van der Vorst, and J. Rommes. *Model Order Reduction: Theory, Research Aspects and Applications*. Springer-Verlag, Berlin/Heidelberg, 2008. [4]

[225] Peter J. Schmid. Nonmodal stability theory. *Annu. Rev. Fluid Mech.*, 39:129–162, 2007. [5]

[226] S. Shank and V. Simoncini. Krylov subspace methods for large scale constrained Sylvester equations. Technical report, Dipartimento di Matematica, Università di Bologna, February 2013. [38]

[227] Fernando C. Silva and Rita Simões. On the Lyapunov and Stein Equations, I. *Linear Algebra and its Applications*, 420(23):329–338, 2007. [21, 33]

[228] Fernando C. Silva and Rita Simões. On the Lyapunov and Stein Equations, II. *Linear Algebra and its Applications*, 426(23):305–311, 2007. [21, 33]

[229] Vasile Sima. *Algorithms for linear-quadratic optimization.* Marcel Dekker, Inc., 1996. [11, 22, 41]

[230] Vasile Sima and Peter Benner. Experimental evaluation of new SLICOT solvers for linear matrix equations based on the matrix sign function. *IEEE Multi-conference on Systems and Control*, pages 601–606, 2008. 9th IEEE International Symposium on Computer-Aided Control Systems Design (CACSD). [30]

[231] V. Simoncini. On the numerical solution of $AX - XB = C$. *BIT*, 36(4):814–830, 1996. [17]

[232] V. Simoncini. A new iterative method for solving large-scale Lyapunov matrix equations. *SIAM J. Sci. Comput.*, 29(3):1268–1288, 2007. [25, 28]

[233] V. Simoncini. The Extended Krylov subspace for parameter dependent systems. *Applied Numer. Math.*, 60(5):550–560, 2010. [15]

[234] V. Simoncini and V. Druskin. Convergence analysis of projection methods for the numerical solution of large Lyapunov equations. *SIAM J. Numer. Anal.*, 47(2):828–843, 2009. [25]

[235] V. Simoncini and M. Sadkane. Arnoldi-Riccati method for large eigenvalue problems. *BIT Numerical Mathematics*, 36(3):579–594, 1996. [5]

[236] V. Simoncini and D. B. Szyld. Recent computational developments in Krylov subspace methods for linear systems. *Num. Lin. Alg. with Appl.*, 14(1):1–59, 2007. [13]

[237] John R. Singler. Approximate low rank solutions of Lyapunov equations via Proper Orthogonal Decomposition. In *2008 American Control Conference, Westin Seattle Hotel, Seattle, Washington, USA*, June 2008. WeA08.3. [31]

[238] Martin Slowik, Peter Benner, and Vasile Sima. Evaluation of the linear matrix equation solvers in SLICOT. *J. of Numerical Analysis, Industrial and Applied Mathematics*, 2(1-2):11–34, 2007. [11, 41]

[239] R. A. Smith. Matrix equation $XA + BX = C$. *SIAM J. Appl. Math.*, 16(1):198–201, 1968. [19, 33]

[240] D. Sorensen and Y. Zhou. Bounds on eigenvalue decay rates and sensitivity of solutions to Lyapunov equations. Technical Report 02-07, Rice University, Houston, Texas, 2002. [21]

[241] D. C. Sorensen and Y. Zhou. Direct methods for matrix Sylvester and Lyapunov equations. *J. Appl. Math.*, 2003:277–303, 2003. [11]

[242] D.C. Sorensen and A.C. Antoulas. The Sylvester equation and approximate balanced reduction. *Lin. Alg. Appl.*, 351–352:671–700, 2002. [5]

[243] D.C. Sorensen and A.C. Antoulas. The Sylvester equation and approximate balanced reduction. *Lin. Alg. Appl.*, 351–352:671–700, 2002. [13]

[244] G. Starke and W. Niethammer. SOR for $AX - XB = C$. *Lin. Alg. Appl.*, 154-156:355–375, 1991. [19]

[245] G. W. Stewart. Error and perturbation bounds for subspaces associated with certain eigenvalue problems. *SIAM Review*, 15:727–764, 1973. [9]

[246] G. W. Stewart and J-G. Sun. *Matrix Perturbation Theory.* Academic Press, 1990. [5, 35]

[247] J. J. Sylvester. Sur l'equations en matrices $px = xq$. *Comptes Rendus Acad. Sci. Paris*, 99:67–71,115–116, 1884. [2]

[248] V.L. Syrmos. Disturbance decoupling using constrained Sylvester equations. *IEEE Transactions on Automatic Control*, 39(4):797–803, 1994. [38]

[249] A. Talbot. The evaluation of integrals of products of linear system responses - part I. *Quart. J. Mech. Appl. Math.*, 12(4):488–503, 1959. [40]

[250] A. Talbot. The evaluation of integrals of products of linear system responses - part II. continued-fraction expansion. *Quart. J. Mech. Appl. Math.*, 12(4):504–520, 1959. [40]

[251] F. De Terán and Froilán M. Dopico. Consistency and efficient solution of the Sylvester equation for $\star$-congruence. *Electronic J. of Linear Algebra*, 22:849–863, 2011. [40]

[252] Fernando De Terán and Froilán M. Dopico. The solution of the equation $XA+AX^T = 0$ and its application to the theory of orbits. *Linear Algebra Appl.*, 434(1):4467, 2011. [41]

[253] Fernando De Terán, Froilán M. Dopico, Nathan Guillery, Daniel Montealegre, and Nicolás Reyes. The solution of the equation $AX + X^\star B = 0$. *Lin. Alg. Appl.*, 438(7):2817–2860, 2013. [40]

[254] Ninoslav Truhar, Zoran Tomljanović, and Ren-Cang Li. Analysis of the solution of the Sylvester equation using low-rank ADI with exact shifts. *Systems & Control Letters*, 59(3-4):248–257, March-April 2010. [10]

[255] Ninoslav Truhar and Krešimir Veselić. Bounds on the trace of a solution to the Lyapunov equation with a general stable matrix. *Systems & Control Letters*, 56(7-8):493–503, July 2007. [27]

[256] C.-C. Tsui. A new approach to robust observer design. *Internat. J. Control*, 47:745–751, 1988. [38]

[257] A. van den Boom, A. Brown, F. Dumortier, A. Geurts, S. Hammarling, R. Kool, M. Vanbegin, P. Van Dooren, and S. Van Huffel. SLICOT, a subroutine library for control and system theory. In *Preprints IFAC Symp. on CADCS*, pages 89–94, Swansea, UK, July 1991. [11, 41]

[258] P. van Dooren. Structured linear algebra problems in digital signal processing. In *Numerical Linear Algebra, Digital Signal Processing and Parallel Algorithms*, NATO Series F, pages 361–384. Springer, 1991. [4]

[259] P. van Dooren and M. Verhaegen. Condensed forms for efficient time-invariant Kalman filtering. *SIAM J. Sci. Statist. Comput.*, 9(3):516–530, May 1988. [38]

[260] Bart Vandereycken. *Riemannian and multilevel optimization for rank-constrained matrix problems*. PhD thesis, Arenberg Doctoral School of Science, Engineering & Technology, Katholieke Universiteit Leuven, December 2010. [24]

[261] Bart Vandereycken and Stefan Vandewalle. A Riemannian optimization approach for computing low-rank solutions of Lyapunov equations. *SIAM J. Matrix Anal. Appl.*, 31(5):2553–2579, 2010. [32]

[262] J.M. Varah. On the separation of two matrices. *SIAM J. Numer. Anal.*, 16(2):216–222, 1979. [9, 16]

[263] Andras Varga. A note on Hammarling's algorithm for the discrete Lyapunov equation. *Systems & Control Letters*, 15(3):273–275, Sept. 1990. [33]

[264] D. Vasilyev and J. White. A more reliable reduction algorithm for behavior model extraction. In *Proceedings of the IEEE/ACM International Conference on Computer-Aided Design*, pages 813–820, 2005. [31]

[265] Dmitry Missiuro Vasilyev. *Theoretical and practical aspects of linear and nonlinear model order reduction techniques*. PhD thesis, Department of Electrical Engineering and Computer Science, MIT, 2008. [5]

[266] Yu. O. Vorontsov and Kh. D. Ikramov. A numerical algorithm for solving the matrix equation $AX + X^T B = C$. *Computational Mathematics and Mathematical Physics*, 51(5):691698, 2011. [40]

[267] K. w. E. Chu. The solution of the matrix equations $AXB - CXD = E$ and $(YA - DZ, YC - BZ) = (E, F)$. *Lin. Alg. Appl.*, 93:93–105, 1987. [35, 37]

[268] E. L. Wachspress. *Iterative solution of elliptic systems*. Prentice-Hall, Inc., Englewood Cliffs, N.J., 1966. [12, 28]

[269] E. L. Wachspress. Iterative Solution of the Lyapunov Matrix Equations. *Appl. Math. Lett.*, 1(1):87–90, 1988. [26]

[270] Eugene L. Wachspress. Extended application of Alternating Direction Implicit iteration model Problem theory. *Journal of the Society for Industrial and Applied Mathematics*, 11(4):994–1016, 1963. [6]

[271] Eugene L. Wachspress. Trail to a Lyapunov equation solver. *Computers and Mathematics with Applications*, 55:1653–1659, 2008. [23]

[272] F. Wan. An in-core finite difference method for separable boundary value problems on a rectangle. *Studies in Appl. Math.*, 52:103–113, 1973. [12]

[273] Qing-Wen Wang, Jian-Hua Sun, and Shang-Zhi Li. Consistency for bi(skew)symmetric solutions to systems of generalized Sylvester equations over a finite central algebra. *Lin. Alg. Appl.*, 353:169182, 2002. [37]

[274] David. S. Watkins. *The matrix eigenvalue problem*. SIAM, Philadelphia, 2007. [12]

[275] H. K. Wimmer. Linear matrix equations, controllability and observability, and the rank of solutions. *SIAM J. Matrix Anal.*, 9(4):570–578, 1988. [8]

[276] H. K. Wimmer. Explicit solutions of the matrix equation $\sum A^i X D_i = C$. *SIAM J. Matrix Anal. Appl.*, 13:1123–1130, 1992. [8]

[277] H. K. Wimmer. The generalized Sylvester equation in polynomial matrices. *IEEE Trans. Automat. Control*, 41:1372–1376, 1996. [36]

[278] H.K. Wimmer. The matrix equation $X - AXB = C$ and an analogue of Roth's theorem. *Lin. Alg. Appl.*, 109:145–147, 1988. [32]

[279] H.K. Wimmer and A.D. Ziebur. Remarks on inertia theorems for matrices. *Czechoslocak Math. J.*, 25:556–561, 1975. [32]

[280] Inc. Wolfram Research. *Mathematica*. Wolfram Research, Inc., Champaign, Illinois, 8.0 edition, 2010. [41]

[281] A.-G. Wu, G.-R. Duan, and B. Zhou. Solution to generalized Sylvester matrix equations. *IEEE Transactions on Automatic Control*, 53(3):811–815, 2008. [37]

[282] A.-G. Wu, J. Hu, and G.-R. Duan. Solutions to the matrix equation $AX-EXF = BY$. *Computers and Mathematics with Applications*, 58(10):1891–1900, 2009. [37]

[283] A.-G. Wu, Y. Sun, and G. Feng. Closed-form solution to the non-homogeneous generalised Sylvester matrix equation. *Control Theory & Applications, IET*, 4(10):1914 – 1921, October 2010. [37]

[284] Ai-Guo Wu, Guang-Ren Duan, and Yu Xue. Kronecker maps and Sylvester-polynomial matrix equations. *IEEE Transactions on Automatic Control*, 52(5), May 2007. [39]

[285] Ai-Guo Wu, Gang Feng, Guang-Ren Duan, and Wei-Jun Wu. Closed-form solutions to Sylvester-conjugate matrix equations. *Computers and Mathematics with Applications*, 60:95–111, 2010. [40]

[286] B. Zhou, G.-R. Duan, and Z.-Y. Li. A Stein matrix equation approach for computing coprime matrix fraction description. *Control Theory & Applications, IET*, 3(6):691–700, 2009. [5]

[287] Bin Zhou and Guang-Ren Duan. Solutions to generalized Sylvester matrix equation by Schur decomposition. *International Journal of Systems Science*, 38(5):369–375, May 2007. [37]

[288] Bin Zhou and Guang-Ren Duan. On equivalence and explicit solutions of a class of matrix equations. *Mathematical and Computer Modelling*, 50:1409–1420, 2009. [37]

[289] Bin Zhou and Guang-Ren Duan. Parametric solutions to the generalized discrete Sylvester matrix equation $MXN - X = TY$ and their applications. *IMA Journal of Mathematical Control and Information*, 26(1):59–78, 2009. [37]

[290] Bin Zhou, Guang-Ren Duan, and Zhao-Yan Li. Gradient based iterative algorithm for solving coupled matrix equations. *Systems & Control Letters*, 58(5):327–333, May 2009. [32, 37]

[291] Bin Zhou, James Lam, and Guang-Ren Duan. On Smith-type iterative algorithms for the Stein matrix equation. *Applied Mathematics Letters*, 22:1038–1044, 2009. [34]

[292] Bin Zhou, James Lam, and Guang-Ren Duan. Toward Solution of Matrix Equation $X = Af(X)B + C$. Technical report, arXiv:1211.0346, 2012. [34]

[293] Yunkai Zhou. *Numerical methods for large scale matrix equations with applications in LTI system model reduction.* PhD thesis, Rice University, Houston, Texas, May 2002. [22, 24]

[294] Yunkai Zhou and D. C. Sorensen. Approximate implicit subspace iteration with alternating directions for LTI system model reduction. *Numer. Linear Algebra Appl.*, 15:873–886, 2008. [5]