ANALYSIS OF KRYLOV SUBSPACE APPROXIMATION TO LARGE-SCALE DIFFERENTIAL RICCATI EQUATIONS[*]

ANTTI KOSKELA[†] AND HERMANN MENA[‡]

Abstract. We consider a Krylov subspace approximation method for the symmetric differential Riccati equation $\dot{X} = AX + XA^T + Q - XSX$, $X(0) = X_0$. The method we consider is based on projecting the large-scale equation onto a Krylov subspace spanned by the matrix A and the low-rank factors of X_0 and Q. We prove that the method is structure preserving in the sense that it preserves two important properties of the exact flow, namely the positivity of the exact flow and also the property of monotonicity. We provide a theoretical a priori error analysis that shows superlinear convergence of the method. Moreover, we derive an a posteriori error estimate that is shown to be effective in numerical examples.

Key words. differential Riccati equations, LQR optimal control problems, large-scale ordinary differential equations, Krylov subspace methods, matrix exponential, exponential integrators, model order reduction, low-rank approximation

AMS subject classifications. 65F10, 65F60, 65L20, 65M22, 93A15, 93C05

1. Introduction. Large-scale differential Riccati equations (DREs) arise in the numerical treatment of optimal control problems governed by partial differential equations. This is the case in particular when solving a linear quadratic regulator problem (LQR), a widely studied problem in control theory. We shortly describe the finite-dimensional LQR problem. For more details, we refer the reader to [1, 8]. The differential Riccati equation arises in the finite horizon case, i.e., when a finite time integral cost functional is considered. Denoting the time interval $[0, t_f]$, the functional has then the quadratic form

$$(1.1) \qquad J(x, u) = \int_0^{t_f} \left(x(t)^T C^T C x(t) + u(t)^T u(t) \right) \mathrm{d}t + x(t_f)^T G x(t_f),$$

where $x(t) \in \mathbb{R}^n$, $C \in \mathbb{R}^{q \times n}$ ($q \ll n$), and $u(t) \in \mathbb{R}^r$ ($r \ll n$). Here $u(t)$ contains the controls, and the matrix C represents the observation matrix. The coefficient matrix G of the penalizing term $x(t_f)^T G x(t_f)$ is symmetric, positive semidefinite and has a low rank. The functional (1.1) is constrained by the system of differential equations

$$(1.2) \qquad \dot{x}(t) = Ax(t) + Bu(t), \qquad x(0) = x_0, \quad t \in [0, t_f],$$

where the matrix $A \in \mathbb{R}^{n \times n}$ is sparse and $B \in \mathbb{R}^{n \times r}$. The number of columns of B corresponds to the number of controls. Under suitable conditions [1, 8], the control $\widetilde{u}$ minimizing the functional (1.1) is given by

$$(1.3) \qquad \widetilde{u}(t) = K(t)\widetilde{x}(t), \quad \text{where} \quad K(t) = -B^T X(t),$$

$X(t)$ is the unique solution of

$$(1.4) \qquad \dot{X} + A^T X + XA - XBB^T X + C^T C = 0, \quad X(t_f) = G,$$

[†]Department of Computer Science, University of Helsinki, Finland (antti.h.koskela@helsinki.fi).
[‡]Department of Mathematics, Yachay Tech, Urcuquí, Ecuador, Department of Mathematics, University of Innsbruck, Austria (mena@yachaytech.edu.ec).

and $\widetilde{x}(t)$ satisfies

$$\dot{\widetilde{x}}(t) = \big(A - BB^T X(t_f - t)\big)\widetilde{x}(t), \quad \widetilde{x}(0) = x_0.$$

As a result, the central computational problem becomes that of solving the final value problem (1.4) which, with a careful change of variables, can be written as an initial value problem.

We consider a Krylov subspace approximation method for large-scale differential Riccati equations of the form (1.4). A projection method for DREs using extended Krylov subspaces has been recently proposed in [15]. We use polynomial Krylov subspaces, and our approach differs from that of [15] in that the initial value matrix G of (1.4) is contained in the Krylov subspace, which allows multiple time stepping. We note that our approach can be also used for extended and rational Krylov subspace methods, and it is related to projection techniques considered for large-scale algebraic Riccati equations [28, 34].

Essentially, the method we consider is based on projecting the matrices A, Q, S, and X_0 on an appropriate Krylov subspace, namely on the *block Krylov subspace* spanned by A and the low-rank factors of X_0 and Q. The projected small-dimensional system is then solved using existing linearization techniques. We show that when using a Padé approximant to solve the linearized small-dimensional system, the total approximation will be structure preserving in a sense that the property of positivity is preserved. Also, the property of monotonicity is preserved under certain conditions. Our Krylov subspace approach is also strongly related to Krylov subspace techniques used for the approximation of the product of a matrix function and a vector, $f(A)b$, and to exponential integrators [21]. For an introduction to matrix functions, we refer the reader to the monograph [18]. The effectiveness of these techniques comes from the fact that generating Krylov subspaces are essentially based on operations of the form $b \rightarrow Ab$, which are cheap for sparse A.

The linearization approach for DREs is a well-known method and allows efficient integration for dense problems; see, e.g., [27]. Another approach, the so called Davison–Maki method [9], uses the fundamental solution of the linearized system. A modified variant, avoiding some numerical instabilities, is proposed in [23]. However, the application of these methods for large-scale problems is impossible due to the high dimensionality of the linearized differential equation.

The problem of solving large-scale DREs has recently received considerable attention. In [4, 5] the authors proposed efficient BDF and Rosenbrock methods for solving DREs capable of exploiting several of the above described properties: sparsity of A, low-rank structure of B, C, and G, and the symmetry of the solution. However, several difficulties arise when approximating the optimal control (1.3) in the large-scale setting. One difficulty is to evaluate the state equation $x(t)$ and Riccati equation $X(t)$ at the same mesh. In [26] a refined ADI integration method is proposed which addresses the high storage requirements of large-scale DRE integrators. In recent studies, an efficient splitting method [36] and adaptive high-order splitting schemes [35] for large-scale DREs have been proposed.

The paper is organized as follows. In Section 2 we describe some preliminaries. Then, in Section 3, the Krylov subspace method is introduced, and its structure preserving properties are shown. In Section 4, the error analysis, first for the differential Lyapunov equation (a simplified version of the DRE) and then for the DRE, is presented. In Section 5 a posteriori error estimation is described. Numerical examples and conclusions in Sections 6 and 7 complete the article.

Notation and definitions. Throughout the article, $\|\cdot\|$ will denote the Euclidean vector norm or its induced matrix norm, i.e., the spectral norm. By $\mathrm{R}(A)$ we denote the column space of a matrix A. To indicate that a matrix A is symmetric positive semidefinite, we write $A \geq 0$. For symmetric matrices A and B we write $B \geq A$ if $B - A \geq 0$.

Also, we will repeatedly use the notion of the *logarithmic norm* of a matrix $A \in \mathbb{C}^{n \times n}$. It can be defined via the *field of values* $\mathcal{F}(A)$, which is defined as

$$\mathcal{F}(A) = \{x^* A x \,:\, x \in \mathbb{C}^n, \|x\| = 1\}.$$

Then, the logarithmic norm $\mu(A)$ of A is defined by

$$\mu(A) \coloneqq \{\max \operatorname{Re} z \,:\, z \in \mathcal{F}(A)\}.$$

We will also repeatedly use the exponential-like function φ_1 defined by

$$\varphi_1(z) = \frac{e^z - 1}{z} = \sum_{\ell=0}^{\infty} \frac{z^\ell}{(\ell+1)!}.$$

2. Preliminaries. From now on we consider the time invariant symmetric differential Riccati equation (DRE) written in the form

(2.1)
$$\dot{X}(t) = AX(t) + X(t)A^T + Q - X(t)SX(t),$$
$$X(0) = X_0,$$

where $t \geq 0$ and $A, Q, S, X_0 \in \mathbb{R}^{n \times n}$, $Q^T = Q$, $S^T = S$, $S \geq 0$. Specifically, we focus on the low-rank symmetric positive semidefinite case, where

$$X_0 = ZZ^T, \quad Q = C^T C,$$

for some $Z \in \mathbb{R}^{n \times p}$ and $C \in \mathbb{R}^{q \times n}$, $p, q \ll n$. Notice that we changed here from A^T to A (a common choice in the numerical analysis literature [10, 11]). Although S arises from the low-rank matrix B in (1.4), we do not place any restriction on the rank of S.

2.1. Linearization. We recall a fact that will be needed later on (see, e.g., [1, Theorem 3.1.1.]).

LEMMA 2.1 (Associated linear system). *The DRE* (2.1) *is equivalent to solving the linear system of differential equations*

(2.2)
$$\frac{\mathrm{d}}{\mathrm{d}t} \begin{bmatrix} U(t) \\ V(t) \end{bmatrix} = \begin{bmatrix} -A & S \\ Q & A^T \end{bmatrix} \begin{bmatrix} U(t) \\ V(t) \end{bmatrix}, \quad \begin{bmatrix} U(0) \\ V(0) \end{bmatrix} = \begin{bmatrix} I \\ X_0 \end{bmatrix},$$

where $U(t), V(t) \in \mathbb{R}^{n \times n}$. If the solution $X(t)$ of (2.1) *exists on the interval $[0, T]$, then the solution of* (2.2) *exists, $U(t)$ is invertible on $[0, T]$, and*

$$X(t) = V(t)U(t)^{-1}.$$

Also, notice that the matrix $\mathcal{H} = \begin{bmatrix} -A & S \\ Q & A^T \end{bmatrix}$ is Hamiltonian, i.e., it holds that

(2.3)
$$(J\mathcal{H})^T = J\mathcal{H}, \quad \text{where} \quad J = \begin{bmatrix} 0 & I \\ -I & 0 \end{bmatrix}.$$

This linearization approach is a standard method for solving finite-dimensional DREs and leads to efficient integration methods for dense problems; see, e.g., [9].

2.2. Integral representation of the exact solution. For the exact solution of (2.1) we have the following integral representation (see also [25, Theorem 8]).

THEOREM 2.2 (Exact solution of the DRE). *The exact solution of the DRE* (2.1) *is given by*

(2.4)
$$X(t) = \mathrm{e}^{tA} X_0 \mathrm{e}^{tA^T} + \int_0^t \mathrm{e}^{(t-s)A} Q \mathrm{e}^{(t-s)A^T} \, \mathrm{d}s$$
$$- \int_0^t \mathrm{e}^{(t-s)A} X(s) S X(s) \mathrm{e}^{(t-s)A^T} \, \mathrm{d}s.$$

Proof. The proof can be carried out by elementary differentiation. □

2.3. Positivity and monotonicity of the exact flow. We recall two important properties of the symmetric DRE, namely the positivity of the exact solution (see, e.g., [10, Prop. 1.1]) and the monotonicity of the solution with respect to the initial data (see, e.g., [11, Theorem 2]). By these we mean the following.

THEOREM 2.3 (Positivity and monotonicity of the solution). *For the solution $X(t)$ of the symmetric DRE* (2.1) *it holds:*

1. *(Positivity) $X(t)$ is symmetric positive semidefinite and exists for all $t > 0$.*
2. *(Monotonicity) Consider two symmetric DREs* (2.1) *corresponding to the linearized systems of the form* (2.2) *with the coefficient matrices*

$$\mathcal{H} = \begin{bmatrix} -A & S \\ Q & A^T \end{bmatrix} \quad and \quad \widetilde{\mathcal{H}} = \begin{bmatrix} -\widetilde{A} & \widetilde{S} \\ \widetilde{Q} & \widetilde{A}^T \end{bmatrix},$$

and let J be the skew-symmetric matrix (2.3). *Then, if $\widetilde{\mathcal{H}}J \leq \mathcal{H}J$ and $0 \leq X_0 \leq \widetilde{X}_0$, it holds that $X(t) \leq \widetilde{X}(t)$ for every $t \geq 0$.*

We will later show that our proposed numerical method preserves the properties of Theorem 2.3.

2.4. Bound for the solution. Using the positivity property of $X(t)$ (Theorem 2.3), we obtain the following bound for the norm of the solution. This will be repeatedly needed in the analysis of the proposed method.

LEMMA 2.4 (Bound for the exact solution). *For the solution $X(t)$ of* (2.1) *it holds that*

$$\|X(t)\| \leq \mathrm{e}^{2t\mu(A)}\|X_0\| + t\varphi_1\big(2t\mu(A)\big)\|Q\|.$$

Proof. Since X_0, Q, and $X(t)$ are all symmetric positive semidefinite, we see that the first two terms on the right-hand side of (2.4) are symmetric positive semidefinite and the third term is symmetric negative semidefinite. Moreover, since $X(t)$ is symmetric positive semidefinite by Theorem 2.3 and since for every symmetric positive definite matrix M it holds that $\|M\| = \max_{\|x\|=1} x^* M x$, we see that

$$\|X(t)\| \leq \left\| \mathrm{e}^{tA} X_0 \, \mathrm{e}^{tA^T} + \int_0^t \mathrm{e}^{(t-s)A} Q \, \mathrm{e}^{(t-s)A^T} \, \mathrm{d}s \right\|.$$

Using the bound $\|\mathrm{e}^{tA}\| \leq \mathrm{e}^{t\mu(A)}$ (see, e.g., [37, p. 138]), the fact that $\mu(A^T) = \mu(A)$, and that $t\varphi_1(tz) = \int_0^t \mathrm{e}^{(t-s)z} \, \mathrm{d}s$, the claim follows. □

From Lemma 2.4 we immediately get the following corollary.

COROLLARY 2.5. *The solution $X(t)$ satisfies*

$$\max_{s \in [0,t]} \|X(s)\| \leq \max\left\{1, e^{2t\mu(A)}\right\} \|X_0\| + t \max\left\{1, \varphi_1\big(2t\mu(A)\big)\right\} \|Q\|.$$

3. A Krylov subspace approximation and its structure preserving properties. In this section we propose our projection method. The original problem (2.1) is projected to a small-dimensional space $R(V_k)$ with V_k having orthonormal columns that span a certain Krylov subspace. The fact that $R(V_k)$ needs to contain this subspace can be seen from the point of view of Krylov subspace approximation of the matrix exponential (see also the solution formula (2.4)). This is strongly related to the approach taken by Saad already in [32] for the algebraic Lyapunov equation. Before introducing our projection method, we recall some basic facts about the Krylov subspace approximation of the matrix exponential. This will also give some auxiliary results that are needed later in the convergence analysis.

3.1. Block Krylov subspace approximation of the matrix exponential. The Krylov subspace approximation of matrix functions has recently been an active topic of research, and we mention the work on classical Krylov subspaces [12, 14, 24, 30], extended Krylov subspaces [24], and rational Krylov subspaces [3, 38].

Block Krylov subspace methods are based on the idea of projecting a high-dimensional problem involving a matrix $A \in \mathbb{R}^{n \times n}$ and a matrix $b \in \mathbb{R}^{n \times \ell}$ onto a lower-dimensional subspace, a block Krylov subspace $\mathcal{K}_k(A, b)$, which is defined by

$$(3.1) \qquad \mathcal{K}_k(A, b) = \text{range}\left[b, Ab, A^2 b, \ldots, A^{k-1} b\right].$$

Usually, an orthogonal basis matrix V_k for $\mathcal{K}_k(A, b)$ is generated using an Arnoldi-type iteration, and this matrix is then used for the projections. There exist several Arnoldi-type methods to produce an orthogonal basis matrix for $\mathcal{K}_k(A, b)$, and in numerical experiments we use the *block Arnoldi iteration* given in [31] which is listed algorithmically as follows:

1. **Input:** $A \in \mathbb{R}^{n \times n}$, $b \in \mathbb{R}^{n \times \ell}$, and the number of iterations k.
2. **Start.** Compute a QR decomposition: $b = U_1 R_1$.
3. **Iterate.** *For $j = 1, ..., k$, compute:*

$$h_{ij} = U_i^T A U_j, \quad i = 1, \ldots, j,$$

$$W_j = A U_j - \sum_{i=1}^{j} U_i h_{ij},$$

$$W_j = U_{j+1} h_{j+1,j} \quad (\textit{QR decomposition of } W_j).$$

As usual, the orthogonalization can be carried out at step 3 as in an element-wise modified Gram–Schmidt manner, and reorthogonalization can be performed if needed. For a robust implementation, deflation techniques could be applied as well; see, e.g., [16].

This algorithm gives a basis matrix with orthogonal columns, $V_k = [U_1 \ \ldots \ U_k] \in \mathbb{R}^{n \times k\ell}$, for the block Krylov subspace $\mathcal{K}_k(A, b)$ and the projected block Hessenberg matrix

$$(3.2) \qquad\qquad\qquad H_k = V_k^T A V_k.$$

This means that the (i, j)-block of size $\ell \times \ell$ of H_k is given by h_{ij} in the above algorithm. Moreover, the following Arnoldi relation holds:

$$A V_k = V_k H_k + U_{k+1} h_{k+1,k} E_k^T,$$

where $E_k = [0 \ \ldots \ 0 \, I_\ell]^T \in \mathbb{R}^{k\ell \times \ell}$.

If A has its field of values on a line, e.g., is Hermitian or skew-Hermitian, then there exists $\theta \in \mathbb{R}$ such that $\mathrm{e}^{\mathrm{i}\theta} A$ is Hermitian. By (3.2) this implies that H_k is block tridiagonal, the orthogonalization recursions become three-term recursions, and we get the *block Lanczos iteration.*

The polynomial approximation property of Krylov subspaces motivates to approximate the product of the matrix exponential e^A and the matrix b as

$$(3.3) \qquad\qquad \mathrm{e}^A b \approx V_k \mathrm{e}^{H_k} V_k^T b = V_k \mathrm{e}^{H_k} E^1 R_1,$$

where $E^1 = [I_\ell \, 0 \ \ldots \ 0]^T \in \mathbb{R}^{k\ell \times \ell}$. For a vector b, the approximation (3.3) was considered originally in [12, 14], and for the case of a block matrix b it has been considered in [29]. For more references for Krylov subspace approximations of matrix functions, see [18, Ch. 13].

Since the columns of V_k are orthonormal, we have $\mathcal{F}(H_k) = \mathcal{F}(V_k^T A V_k) \subset \mathcal{F}(A)$, and from this it follows that $\mu(H_k) \le \mu(A)$. Clearly, it also holds that $\|H_k\| \le \|A\|$. Moreover, we have the following bound:

LEMMA 3.1. *For the approximation* (3.3)*, it holds that*

$$\|\mathrm{e}^{tA} b - V_k \mathrm{e}^{tH_k} V_k^T b\| \le 2 \max\left\{1, \mathrm{e}^{t\mu(A)}\right\} \frac{\|tA\|^k}{k!} \|b\|.$$

Proof. The proof goes analogously to the proof of [14, Theorem 2.1], where b is a vector. □

3.2. Rational Krylov subspaces. We also mention the possibility of approximating matrix functions in *rational Krylov subspaces* ; see, e.g., [13, 17, 34, 38].

For poles $\bar{s} = \{s_1, s_2, \ldots\}$, $s_i \in \mathbb{C}$, the rational Krylov subspace can be defined as

$$(3.4) \qquad \mathcal{K}_k(A, b, \bar{s}) = \mathrm{span}\{b, (s_1 I - A)^{-1} b, \ldots, \prod_{\ell=1}^{k-1} (s_\ell I - A)^{-1} b\}.$$

Then, if a matrix V_k with orthogonal columns gives a basis for the subspace $\mathcal{K}_k(A, b, \bar{s})$, the matrix exponential can be approximated as (3.3), where $H_k = V_k^T A V_k$. Especially for sparse matrices, the rational Krylov methods are often more efficient, and as the solution usually converges faster with respect to the subspace dimension, the rational alternative is usually more memory efficient. These differences will be illustrated in numerical experiments. However, for simplicity, in our analysis and numerical experiments we will use the polynomial Krylov subspace method.

3.3. The method. We approximate $X(t)$ in the block Krylov subspace $\mathcal{K}_k\big(A, \big[Z \ \ C^T\big]\big)$. The fact that the projection onto this subspace results in an accurate approximation can be seen from the form of the exact solution (2.4) and from the Krylov approximation properties shown in the last section. To this end, an orthogonal basis matrix $V_k \in \mathbb{R}^{n \times k(p+q)}$ for $\mathcal{K}_k\big(A, \big[Z \ \ C^T\big]\big)$ is first generated using the block Arnoldi iteration. Then, we project the problem (2.1) using V_k and solve the projected differential equation to obtain the solution $Y_k(t) \in \mathbb{R}^{k \times k}$. Then, we project back to $\mathbb{R}^{n \times n}$ using V_k to obtain the numerical solution $X_k(t) := V_k Y_k(t) V_k^T$. This procedure is listed in Algorithm 1. Notice that there is no restriction on the rank of S.

In practice, multiple time stepping is often needed, and in that case Algorithm 1 can be used for approximating a single time step. We also note that then the large-dimensional matrix $X_k(t)$ should not be explicitly formed, but the matrices V_k and $Y_k(t)$ should only be used for constructing the initial value for the next time step. The a posteriori estimates of Section 5 can be added to Algorithm 1 to decide whether the Krylov subspace size k is sufficiently large.

Algorithm 1: Krylov subspace approximation of the DRE (2.1).

Input: Krylov subspace size k, matrices $A, S \in \mathbb{R}^{n \times n}$, $Z \in \mathbb{R}^{n \times p}$, and $C \in \mathbb{R}^{q \times n}$,
where $p, q \ll n$.

1 Carry out k steps of the block Arnoldi iteration to obtain
- the orthogonal basis matrix V_k of $\mathcal{K}_k\big(A, \begin{bmatrix} Z & C^T \end{bmatrix}\big)$,
- the block Hessenberg matrix $H_k = V_k^T A V_k$,
- and the matrices $C_k = C V_k$ and $Z_k = V_k^T Z$.

2 Compute $S_k = V_k^T S V_k$.

3 Compute the solution $Y_k(t)$ for the small-dimensional system

$$(3.5) \qquad \begin{aligned} \dot{Y}_k(t) &= H_k Y_k(t) + Y_k(t) H_k^T + C_k^T C_k - Y_k(t) S_k Y_k(t), \\ Y_k(0) &= Z_k Z_k^T \end{aligned}$$

using the modified Davison–Maki method.

4 Approximate $X(t) \approx X_k(t) = V_k Y_k(t) V_k^T$.

3.4. Solving the small-dimensional system. In the numerical implementation we use the *modified Davison–Maki method* [23] to solve the small-dimensional system (3.5). This method is chosen because of its structure preserving properties, which are shown in Section 3.5. The method can be described as follows.

As shown in Lemma 2.1, the solution of the projected system (3.5) is given by

$$(3.6) \quad Y_k(t) = W_k(t) U_k(t)^{-1}, \text{ where } \quad \begin{bmatrix} U_k(t) \\ W_k(t) \end{bmatrix} = \exp\left(t \begin{bmatrix} -H_k & S_k \\ C_k^T C_k & H_k^T \end{bmatrix} \right) \begin{bmatrix} I_k \\ Z_k Z_k^T \end{bmatrix}.$$

Instead of directly evaluating $Y_k(t)$ by (3.6), which is the idea of the original Davison–Maki method [9], we perform substepping in order to avoid numerical instabilities arising from the inversion of the matrix $U_k(t)$ in (3.6). This is exactly the modified Davison–Maki method, and it is presented by the following pseudocode. We set $Y_k^j \approx Y_k\big(\frac{j \cdot t}{m}\big)$.

1. **Input:** Hamiltonian matrix $\begin{bmatrix} -H_k & S_k \\ C_k^T C_k & H_k^T \end{bmatrix}$, $Y_k(0) = Z_k Z_k^T$,
   time $t > 0$, substep size $\Delta t = t/m$, $m \in \mathbb{Z}_+$.

2. **Set:** $Y_k^0 = Y_k(0)$.

3. **Iterate.** *For $j = 0, ..., m - 1$:*

$$\begin{bmatrix} U_{j+1} \\ W_{j+1} \end{bmatrix} = \exp\left(\Delta t \begin{bmatrix} -H_k & S_k \\ C_k^T C_k & H_k^T \end{bmatrix} \right) \begin{bmatrix} I_k \\ Y_k^j \end{bmatrix}, \qquad Y_k^{j+1} = W_{j+1} U_{j+1}^{-1}.$$

For computing the matrix exponential in Step 3, we use the $13th$ order diagonal Padé aproximant which is implemented in Matlab as the $\texttt{expm}$ command [19]. We note that this approach may become computationally infeasible for large values of p and q and/or for a large Krylov subspace size k as the matrix exponential has size $2k(p + q) \times 2k(p + q)$. For solving the system (3.5) also other methods designed for DREs could be used; see, e.g., [4] and [5].

3.5. Structure preserving properties of the approximation. We next inspect the two properties stated in Theorem 2.3. We show that the proposed method preserves the property of positivity of the exact flow, and it also preserves the property of monotonicity under the condition that the matrix V_k stays constant when the initial data for the DRE is changed. Notice that these results are not restricted to polynomial Krylov subspace methods.

THEOREM 3.2. *The numerical approximation given by Algorithm 1 together with the modified Davison–Maki method for the small-dimensional system preserves the property of positivity stated in Theorem 2.3.*

Proof. The projected coefficient matrices S_k, $C_k^T C_k$ and the initial value $Z_k Z_k^T$ of the small system (3.5) are clearly all symmetric positive semidefinite. Thus, the small system (3.5) is a symmetric DRE. By Theorem 3.1 of [10], an application of a symplectic Runge–Kutta scheme with positive weights b_i (see [11] for details) gives a symmetric positive semidefinite solution. As the sth-order diagonal Padé approximant equals the stability function of the s-stage Gauss–Legendre method (see, e.g., [22, p. 46]), the Padé approximation in the third substep of the modified Davison–Maki method (Section 3.4) corresponds to a symplectic Runge–Kutta method. Thus, each substep of the modified Davison–Maki method outputs a symmetric positive semidefinite matrix, and as a result $Y_k(t)$ is symmetric positive semidefinite. Therefore also $X_k(t) = V_k Y_k(t) V_k^T$ is symmetric positive semidefinite.   □

THEOREM 3.3. *The numerical approximation given by Algorithm 1 together with the modified Davison–Maki method for the small-dimensional system preserves the property of monotonicity in the following sense. Consider two DREs corresponding to linearizations with the coefficient matrices*

$$(3.7) \qquad \mathcal{H} = \begin{bmatrix} -A & S \\ Q & A^T \end{bmatrix} \quad and \quad \widetilde{\mathcal{H}} = \begin{bmatrix} -\widetilde{A} & \widetilde{S} \\ \widetilde{Q} & \widetilde{A}^T \end{bmatrix}$$

such that

$$(3.8) \qquad \widetilde{\mathcal{H}}J \le \mathcal{H}J, \quad 0 \le X_0 \le \widetilde{X}_0.$$

Suppose both systems are projected using the same orthogonal basis matrix $V_k \in \mathbb{R}^{n \times k}$, giving as a result small k-dimensional systems of the form (3.5) for the matrices $Y_k(t)$ and $\widetilde{Y}_k(t)$. Then, for the matrices $X_k(t) = V_k Y_k(t) V_k^T$ and $\widetilde{X}_k(t) = V_k \widetilde{Y}_k(t) V_k^T$ we have

$$X_k(t) \le \widetilde{X}_k(t).$$

Proof. Consider the projected systems of the form (3.5) corresponding to $Y_k(t)$ and $\widetilde{Y}_k(t)$ with the projected coefficient matrices H_k, Q_k, and S_k, and $\widetilde{H}_k$, $\widetilde{Q}_k$, and $\widetilde{S}_k$, respectively. Consider also the corresponding linearizations of the form (3.7) with the Hamiltonian matrices

$$\mathcal{H}_k := \begin{bmatrix} V_k & 0 \\ 0 & V_k \end{bmatrix}^T \mathcal{H} \begin{bmatrix} V_k & 0 \\ 0 & V_k \end{bmatrix} \quad and \quad \widetilde{\mathcal{H}}_k := \begin{bmatrix} V_k & 0 \\ 0 & V_k \end{bmatrix}^T \widetilde{\mathcal{H}} \begin{bmatrix} V_k & 0 \\ 0 & V_k \end{bmatrix}.$$

By the reasoning of the proof of Theorem 3.2, the projected systems corresponding to $Y_k(t)$ and $\widetilde{Y}_k(t)$ are symmetric DREs. We see that

$$\widetilde{\mathcal{H}}_k J_k - \mathcal{H}_k J_k = \begin{bmatrix} V_k & 0 \\ 0 & V_k \end{bmatrix}^T (\widetilde{\mathcal{H}}J - \mathcal{H}J) \begin{bmatrix} V_k & 0 \\ 0 & V_k \end{bmatrix},$$

where $J_k = \begin{bmatrix} 0 & I \\ -I & 0 \end{bmatrix} \in \mathbb{R}^{2k \times 2k}$. Thus, from (3.8) it follows that $\widetilde{\mathcal{H}}_k J_k \le \widetilde{\mathcal{H}}_k J_k$. Clearly, also $0 \le Y_k(0) \le \widetilde{Y}_k(0)$. By [11, Theorem 6], an application of a symplectic Runge–Kutta scheme

with positive weights b_i (see [11] for details) preserves the monotonicity. Thus, the Padé approximants of the substeps of the modified Davison–Maki method (Section 3.4) preserve the monotonicity. Therefore, $Y_k(t) \leq \widetilde{Y}_k(t)$, and as a consequence $X_k(t) \leq \widetilde{X}_k(t)$.   □

REMARK 3.4. As the basis matrix V_k given by Algorithm 1 is independent of the matrix $S = BB^T$ in the DRE (2.1), where B is the control matrix in the original linear system (1.2), we see that Algorithm 1 preserves monotonicity under modifications of B. However, if we change the initial value X_0 or the matrix Q, then forming a new basis V_k is generally needed. The fact that V_k is independent of B can also be seen by considering similar projection methods for the algebraic Riccati equation; see, e.g., [33] and the references therein.

4. A priori error analysis. We first consider the approximation of the DRE without the quadratic term $-XSX$, i.e., we consider the differential Lyapunov equation. This clarifies the presentation as the derived bounds will be needed when we consider the approximation of the differential Riccati equation. We note, however, that the bounds for the Lyapunov equation are applicable outside of the scope of the optimal control problems, e.g., when considering time integration of an inhomogeneous matrix differential equation.

4.1. Error analysis for the Lyapunov equation. Consider the symmetric Lyapunov differential equation with low-rank initial data and constant low-rank inhomogeneity,

$$(4.1) \qquad \begin{aligned} \dot{X}(t) &= AX(t) + X(t)A^T + C^T C, \\ X(0) &= ZZ^T, \end{aligned}$$

where $Z \in \mathbb{R}^{n \times p}$ and $C \in \mathbb{R}^{q \times n}$, with $p, q \ll n$. Then, the approximation is given by $X_k(t) = V_k Y_k(t) V_k^T$, where $Y_k(t)$ is a solution of the projected system (3.5) with $S = 0$. For the error of this approximation we obtain the following bound.

THEOREM 4.1. *Let $A \in \mathbb{R}^{n \times n}$, $Z \in \mathbb{R}^{n \times p}$, $C \in \mathbb{R}^{q \times n}$, and let $X(t)$ be the solution of (4.1). Let $V_k \in \mathbb{R}^{n \times m(q+p)}$ be an orthogonal basis for the block Krylov subspace $\mathcal{K}_k\big(A, \begin{bmatrix} Z & C^T \end{bmatrix}\big)$. Let $Y_k(t)$ be the solution of the projected system (3.5) with $S = 0$, and let $X_k(t) = V_k Y_k(t) V_k^T$. Then,*

$$\|X(t) - X_k(t)\| \leq 4 \max\left\{1, \mathrm{e}^{2t\mu(A)}\right\} \|A\|^k \left(\frac{t^k}{k!} \|X_0\| + \frac{t^{k+1}}{(k+1)!} \|Q\| \right).$$

Proof. Using the integral representation of Theorem 2.2 for both $X(t)$ and $Y_k(t)$, we see that

$$X(t) - X_k(t) = Err_{1,k}(t) + Err_{2,k}(t),$$

where

$$(4.2) \qquad Err_{1,k}(t) = \mathrm{e}^{tA} ZZ^T \mathrm{e}^{tA^T} - V_k \mathrm{e}^{tH_k} V_k^T ZZ^T V_k \mathrm{e}^{tH_k^T} V_k^T,$$

and

$$(4.3) \qquad \begin{aligned} Err_{2,k}(t) &= \int_0^t \mathrm{e}^{(t-s)A} C^T C \mathrm{e}^{(t-s)A^T} \, \mathrm{d}s \\ &\quad - \int_0^t V_k \mathrm{e}^{(t-s)H_k} V_k^T C^T C V_k \mathrm{e}^{(t-s)H_k^T} V_k^T \, \mathrm{d}s. \end{aligned}$$

Adding and subtracting $\mathrm{e}^{tA}ZZ^T V_k \mathrm{e}^{tH_k^T} V_k^T$ on the right-hand side of (4.2) gives

$$Err_1(t) = \mathrm{e}^{tA}Z \left(\mathrm{e}^{tA}Z - V_k \mathrm{e}^{tH_k} V_k^T Z\right)^T$$
$$- \left(V_k \mathrm{e}^{tH_k} V_k^T Z - \mathrm{e}^{tA}Z\right) Z^T V_k \mathrm{e}^{tH_k^T} V_k^T.$$

Using Lemma 3.1 to bound the norm of $\mathrm{e}^{tA}Z - V_k \mathrm{e}^{tH_k} V_k^T Z$ and the facts that $\mu(H_k) \leq \mu(A)$ and that $\|X_0\| = \|ZZ^T\| = \|Z\|^2$ gives

$$\|Err_1(t)\| \leq 4 \max\left\{1, \mathrm{e}^{2t\mu(A)}\right\} \frac{\|tA\|^k}{k!} \|X_0\|.$$

Then, similarly, adding and subtracting the term $\int_0^t \mathrm{e}^{(t-s)A} C^T C V_k \mathrm{e}^{(t-s)H_k^T} V_k^T \, \mathrm{d}s$ to (4.3) and applying Lemma 3.1 shows that

$$\|Err_2(t)\| \leq 4\|Q\| \int\limits_0^t \max\left\{1, \mathrm{e}^{2(t-s)\mu(A)}\right\} \frac{\|(t-s)A\|^k}{k!} \, \mathrm{d}s$$

$$\leq 4\|Q\| \max\left\{1, \mathrm{e}^{2t\mu(A)}\right\} \|A\|^k \frac{t^{k+1}}{(k+1)!},$$

which shows the claim.   ☐

We note that the error bound of Theorem 4.1, similarly to the bounds given in [14], exhibits a hump before it starts to decrease in case $\|tA\| > 1$.

4.2. Refined error bounds for the Lyapunov equation. Although Theorem 4.1 shows the superlinear convergence for the approximation of the Lyapunov equation (4.1), sharper bounds can be obtained, e.g., by using the bounds given in [20]. As an example we consider the following. If A is symmetric negative semidefinite with its spectrum inside the interval $[-4\rho, 0]$ and V_k is an orthogonal basis matrix for the block Krylov subspace $\mathcal{K}_k(A, B)$, then we have (see [20, Theorem 2]) the bound for the error $\varepsilon_k := \|\mathrm{e}^{tA}B - V_k \mathrm{e}^{tH_k} V_k^T B\|$

(4.4)
$$\varepsilon_k \leq 10\, \mathrm{e}^{-k^2/(5\rho t)}\|B\|, \qquad \sqrt{4\rho t} \leq k \leq 2\rho t,$$

$$\varepsilon_k \leq 10\, (\rho t)^{-1}\mathrm{e}^{-\rho t} \left(\frac{\mathrm{e}\rho t}{k}\right)^k, \qquad k \geq 2\rho t.$$

Using (4.4) and following the proof of Theorem 4.1, we get the following bound for the case of a symmetric negative semidefinite A.

THEOREM 4.2. *Let* $A \in \mathbb{R}^{n \times n}$, $Z \in \mathbb{R}^{n \times p}$, *and let* $C \in \mathbb{R}^{q \times n}$ *define the Lyapunov equation* (4.1). *Let* $V_k \in \mathbb{R}^{n \times m(q+p)}$ *be an orthogonal basis matrix for the subspace* $\mathcal{K}_k(A, \begin{bmatrix} Z & C^T \end{bmatrix})$. *Let* $Y_k(t)$ *be the solution of the projected (using* V_k*) system* (3.5) *with* $S = 0$, *and let* $X_k(t) = V_k Y_k(t) V_k^T$. *Then, for the error* $\varepsilon_k := \|X(t) - X_k(t)\|$ *it holds that*

$$\varepsilon_k \leq 20\, \mathrm{e}^{-k^2/(5\rho t)}\big(\|X_0\| + t\|Q\|\big), \qquad \sqrt{4\rho t} \leq k \leq 2\rho t,$$

$$\varepsilon_k \leq 20\, (\rho t)^{-1}\mathrm{e}^{-\rho t} \left(\frac{\mathrm{e}\rho t}{k}\right)^k \big(\|X_0\| + t\|Q\|\big), \qquad k \geq 2\rho t.$$

4.3. Error for the approximation of the Riccati equation. Here, we state our main theorem which shows the superlinear convergence property of Algorithm 1 when applied to the DRE (2.1). Its proof, which is essentially based on Lemma 3.1 and Grönwall's lemma, is lengthy and is left to Appendix A.

First, however, we state a bound for the norm of the numerical solution $X_k(t)$, which will be needed in the proof of the main theorem.

LEMMA 4.3. *Suppose that $X_0 = ZZ^T$, $Q = C^T C$, and that S is symmetric positive semidefinite. Then, $X_k(t)$ is symmetric positive semidefinite and satisfies the bound*

$$\|X_k(t)\| \le \mathrm{e}^{2t\mu(A)}\|X_0\| + t\varphi_1\big(2t\mu(A)\big)\|Q\|.$$

Proof. As ZZ^T, $C^T C$, and S are symmetric and positive semidefinite, we see from (3.5) that so are the orthogonally projected matrices $Z_k Z_k^T$, $C_k^T C_k$, and S_k. Thus, the projected system is a symmetric DRE. Applying Lemma 2.4 to the projected system and using the bounds $\mu(H_k) \le \mu(A)$, $\|Q_k\| \le \|Q\|$, and $\|V_k V_k^T X_0 V_k V_k^T\| \le \|X_0\|$ shows the claim.   □

From Lemma 4.3 we immediately get the following bound.

COROLLARY 4.4. *The numerical solution $X_k(t)$ satisfies*

$$\max_{s \in [0,t]} \|X_k(s)\| \le \max\left\{1, \mathrm{e}^{2t\mu(A)}\right\}\|X_0\| + t\max\left\{1, \varphi_1\big(2t\mu(A)\big)\right\}\|Q\|.$$

We are now ready to state an error bound for the DRE. The proof is left to Appendix A.

THEOREM 4.5. *Let $A \in \mathbb{R}^{n \times n}$, $Z \in \mathbb{R}^{n \times p}$, $C \in \mathbb{R}^{q \times n}$, and let $S \in \mathbb{R}^{n \times n}$ define the DRE (2.1). Let $X_k(t)$ be the numerical solution given by Algorithm 1. Then, the following bound holds:*

$$\|X(t) - X_k(t)\| \le c(t)\|A\|^k \left(\frac{t^k}{k!}\|X_0\| + \frac{t^{k+1}}{(k+1)!}\|Q\|\right),$$

where

$$c(t) = 4\left(1 + 2\|S\|\alpha(t)\max\left\{1, \mathrm{e}^{t\mu(A)}\right\}c_2(t)\right)\mathrm{e}^{t\|S\|\alpha(t)},$$

$$c_2(t) = 1 + t\|S\|\alpha(t)\varphi_1\left(t\|S\|\alpha(t)\max\left\{1, \mathrm{e}^{t\mu(A)}\right\}\right),$$

$$\alpha(t) = \max\left\{1, \mathrm{e}^{2t\mu(A)}\right\}\|X_0\| + t\max\left\{1, \varphi_1\big(2t\mu(A)\big)\right\}\|Q\|.$$

5. Heuristic a posteriori error estimate. We consider next an a posteriori error estimation for the method by using ideas presented in [7].

Denote the original DRE (2.1) as

$$\dot{X}(t) = F(X(t)), \quad X(0) = X_0.$$

Using the residual matrix $R_k(t) = F(X_k(t)) - \dot{X}_k(t)$, we derive computable error estimates. These derivations are based on the following lemma.

LEMMA 5.1. *The error $\mathcal{E}_k(t) \coloneqq X(t) - X_k(t)$ satisfies the equation*

(5.1)
$$\mathcal{E}_k(t) = \int_0^t \mathrm{e}^{(t-s)A} R_k(s)\mathrm{e}^{(t-s)A^T}\,\mathrm{d}s$$
$$- \int_0^t \mathrm{e}^{(t-s)A}\Big(\mathcal{E}_k(s)SX(s) + X_k(s)S\mathcal{E}_k(s)\Big)\mathrm{e}^{(t-s)A^T}\,\mathrm{d}s,$$

where

(5.2)
$$R_k(t) = U_{k+1}H_{k+1,k}E_k^T Y_k(t)V_k^T + V_k Y_k(t)E_k H_{k+1,k}^T U_{k+1}^T.$$

Proof. We see that the error $\mathcal{E}_k(t)$ satisfies the ODE

(5.3)
$$
\begin{aligned}
\dot{\mathcal{E}}_k(t) = \dot{X}(t) - \dot{X}_k(t) &= F(X(t)) - F(X_k(t)) + R_k(t) \\
&= A\big(X(t) - X_k(t)\big) + \big(X(t) - X_k(t)\big)A^T \\
&\quad - X(t)SX(t) + X_k(t)SX_k(t) + R_k(t) \\
&= A\big(X(t) - X_k(t)\big) + \big(X(t) - X_k(t)\big)A^T \\
&\quad - \big(X(t) - X_k(t)\big)SX(t) - X_k(t)S\big(X(t) - X_k(t)\big) + R_k(t) \\
&= A\mathcal{E}_k(t) + \mathcal{E}_k(t)A^T - \mathcal{E}_k(t)SX(t) - X_k(t)S\mathcal{E}_k(t) + R_k(t),
\end{aligned}
$$

with the initial value $\mathcal{E}_k(0) = 0$. Applying the variation-of-constants formula to (5.3) gives (5.1).

Next we show the representation (5.2). Since

$$
F(X_k(t)) = AV_kY_k(t)V_k^T + V_kY_k(t)V_k^TA^T + Q - V_kY_k(t)V_k^TSV_kY_k(t)V_k^T
$$

and

$$
\dot{X}_k(t) = V_kH_kY_k(t)V_k^T + V_kY_k(t)H_k^TV_k^T + V_kQ_kV_k^T - V_kY_k(t)V_k^TSV_kY_k(t)V_k^T,
$$

we see that

(5.4)
$$
\begin{aligned}
R_k(t) &= F(X_k(t)) - \dot{X}_k(t) \\
&= (AV_k - V_kH_k)Y_k(t)V_k^T + V_kY_k(t)(AV_k - V_kH_k)^T + Q - V_kQ_kV_k^T \\
&= (AV_k - V_kH_k)Y_k(t)V_k^T + V_kY_k(t)(AV_k - V_kH_k)^T,
\end{aligned}
$$

since $V_kQ_kV_k^T = V_kV_k^TC^TCV_kV_k^T = C^TC = Q$ as $C^T \in \mathrm{R}(V_k)$. Substituting the Arnoldi relation $AV_k - V_kH_k = U_{k+1}H_{k+1,k}E_k^T$ into (5.4) gives the representation (5.2). $\square$

To derive a heuristic a posteriori estimate, we neglect the second term in equation (5.1) and approximate the first integral by leaving out the exponentials. This leads to the approximation

(5.5)
$$
\begin{aligned}
\mathcal{E}_k(t) \approx \int_0^t R_k(s)\,\mathrm{d}s &= U_{k+1}H_{k+1,k}E_k^T\Big(\int_0^t Y_k(s)\,\mathrm{d}s\Big)V_k^T \\
&\quad + V_k\Big(\int_0^t Y_k(s)\,\mathrm{d}s\Big)^TE_kH_{k+1,k}^TU_{k+1}^T.
\end{aligned}
$$

From a careful inspection we see that $U_{k+1}^TV_k = 0$ implies

(5.6)
$$
\begin{aligned}
\Big\|\int_0^t R_k(s)\,\mathrm{d}s\Big\| &= \Big\|U_{k+1}H_{k+1,k}E_k^T\Big(\int_0^t Y_k(s)\,\mathrm{d}s\Big)V_k^T\Big\| \\
&= \Big\|H_{k+1,k}E_k^T\Big(\int_0^t Y_k(s)\,\mathrm{d}s\Big)\Big\|.
\end{aligned}
$$

The integral $\int_0^t Y_k(s)\,\mathrm{d}s$ can be estimated by a simply quadrature

$$
\int_0^t Y_k(s)\,\mathrm{d}s \approx \sum_{\ell=1}^m \Delta tY_k(\ell\Delta t),
$$

where $\Delta t = t/m$. The intermediate values $Y_k(\ell\Delta t)$ can be obtained directly from the summing and squaring phase of the modified Davison–Maki method (Section 3.4). From (5.5) and (5.6), we arrive at a computationally efficient a posteriori estimate

$$(5.7) \qquad est_k := \left\| H_{k+1,k} E_k^T \sum_{\ell=1}^m \Delta t Y_k(\ell\Delta t) \right\|.$$

To illustrate the efficiency of this estimate consider the following example. Let $A \in \mathbb{R}^{400\times400}$ be the tridiagonal matrix $10^2 \cdot \mathrm{diag}(1,-2,1)$, $t = 0.1$, and let $Z \in \mathbb{R}^{400}$, $C \in \mathbb{R}^{400}$, and $B \in \mathbb{R}^{400}$ $(S = BB^T)$ be random vectors. Figure 5.1 shows the error $\|X(t) - X_k(t)\|$ versus the estimate (5.7). The reference solution is computed using a Matlab ODE solver with a small error tolerance.
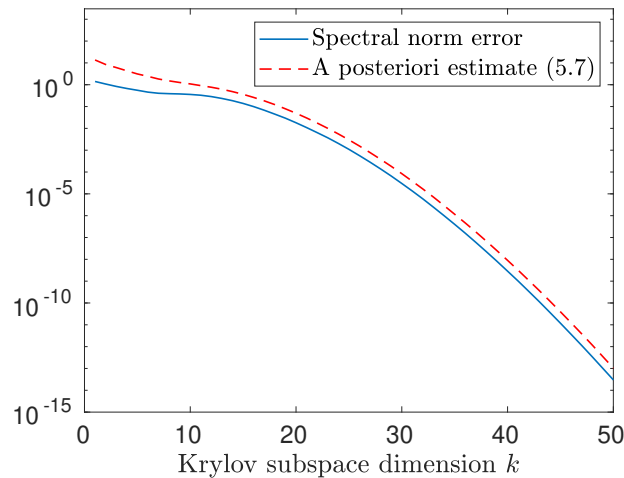

FIGURE 5.1. *Convergence of the approximation versus the a posteriori estimate* (5.7).

We note that by using the error representation (5.1) and the residual $R_k(t)$ given in (5.2), it is possible to derive corrected schemes, similarly as is done for the matrix exponential in [7] and [30].

6. Numerical experiments: optimal cooling problem. As a numerical example we consider an optimal cooling problem described in [6] (see also Example 2 in [36]). The underlying linear system is of the form

$$(6.1) \qquad \begin{aligned} M\dot{x} &= Ax + Bu, \\ y &= Cx, \end{aligned}$$

where the coefficient matrices arise from a finite element discretization of the cross section of a rail. A discretization of dimension n gives coefficient matrices $A, M \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times7}$, and $C \in \mathbb{R}^{6\times n}$, where A is symmetric. This leads to a symmetric DRE of the form (2.1) with the coefficient matrices $\widetilde{A} = M^{-1}A$, $Q = C^TC$, and $S = M^{-1}B(M^{-1}B)^T$. We take zero initial values for the DRE. The mass matrix M is sparse so the products using the matrix $M^{-1}A$ are cheap. We note that by a symmetric decomposition $M = L^TL$, the system (6.1) can also be written as a linear system for the scaled variable Lx using the coefficient matrices

$\widetilde{A} = L^{-T}AL^{-1}$, $\widetilde{B} = L^{-T}B$, and $\widetilde{C} = CL^{-1}$. This leads to a symmetric DRE with a symmetric coefficient matrix $\widetilde{A}$.

We emphasize that the numerical experiments are only to illustrate the theoretical results and therefore the dimensions ($n = 1357$ and $n = 5177$) of the benchmark cases are rather small. Moreover, we note that the considered algorithm is not competitive to methods based on more efficient basis choices such as rational Krylov subspace methods.

6.1. Case $n = 1357$. Figure 6.1 displays the convergence of Algorithm 1 and the a posteriori error estimate given by (5.7) when $T = 10$. We compute the spectral norm error $\|X(T) - X_k(T)\|$ for different Krylov subspace dimensions k. For the scaling and squaring part (Section 3.4), we set the parameter $m = 10$.

Figure 6.2 illustrates the convergence of a single step for $T = 20$ when we apply the block orthogonalization procedure of Section 3.1 for the Krylov subspace (3.1) and for a rational Krylov subspace (3.4) spanned by A. For the rational Krylov subspace we set all nodes s_i equal to 1; see [17, Example 3.5]. We emphasize that this is a non-optimal choice of nodes and that by careful selection, better convergence would be achieved [17]. Here, the subspace dimension denotes the number of columns of the basis matrix V_k. For comparison, we also consider the best low-rank approximation of the solution $X(T)$ obtained from its singular value decomposition (SVD) for different ranks (denoted as basis dimension in the figure). We observe that the rational approximation needs a considerably smaller subspace for a given error than the polynomial approximation.
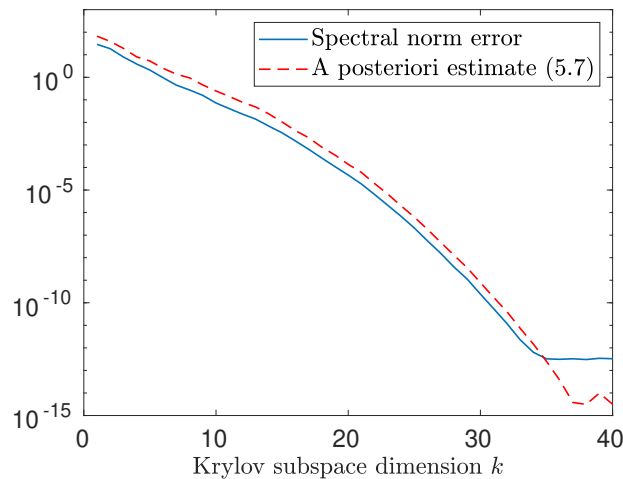

FIGURE 6.1. *Convergence of the approximation given by Algorithm 1 and its a posteriori estimate* (5.7).

Next, we apply Algorithm 1 for $N = 10$ subsequent time steps. We set for the Krylov error a tolerance ε and use the a posteriori estimate (5.7) as a criterion for stopping the iteration. Also, after each step we cut the rank using the projector P_ε defined for a matrix X with singular triplets (σ_i, u_i, v_i) by

$$P_\epsilon(X) := \sum_{\sigma_i > \epsilon} \sigma_i u_i v_i^T.$$

Figure 6.3 depicts the final errors at $T = 10$ for 4 different values of ε. As we see, the final errors are not far from the tolerances ε used for substeps. Figure 6.4 depicts the growth of
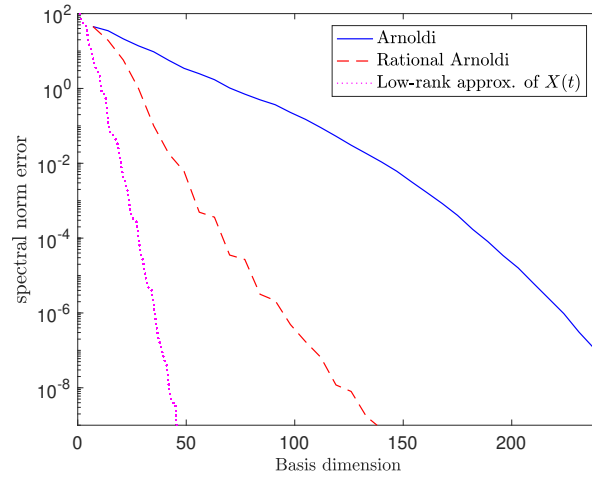
FIGURE 6.2. *Convergence of a single step approximation with the Arnoldi and the rational Arnoldi iteration. The figure displays also the convergence of the best low-rank approximation of $X(t)$.*
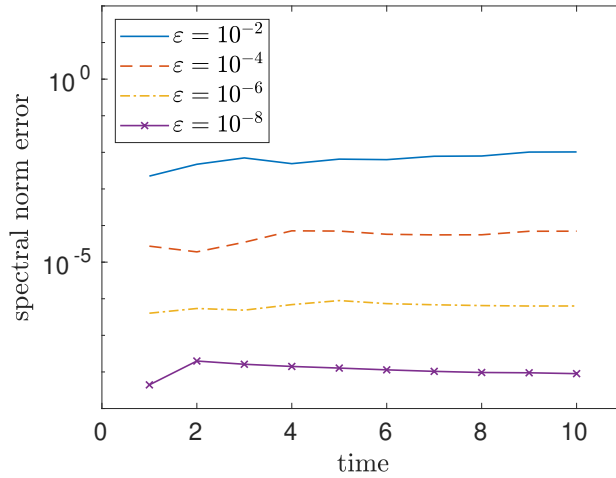

FIGURE 6.3. *The error of the numerical solution for different tolerances ε.*

the rank in the numerical solution for different tolerances ε. We observe that the substepping approach requires less memory for a given error tolerance than a single run using Algorithm 1. This is depicted in Table 6.1. Notice that the difference in the numbers of Table 6.1 and those of Figure 6.4 come from the fact that the Krylov iteration multiplies the memory needed for the basis vectors. This difference can be reduced by the use of rational Krylov methods.

As a last experiment for the case $n = 1357$, we carry out time integration up to $T = 4500$ using $N = 900$ time steps. We use a Krylov subspace dimension $k = 32$ for the first time step and $k = 20$ for the rest. After each time step, we cut the rank of the numerical solution to 40 using the SVD. By these choices of Krylov subspace sizes, the error arising from the rank cut dominates at each time step. Assuming that the error arising from the rank cut is much larger
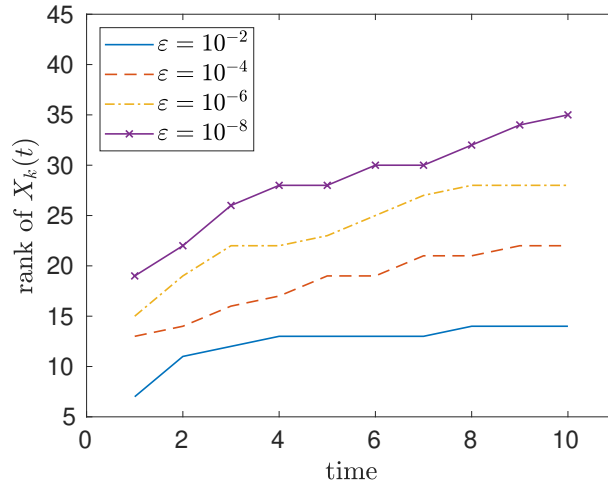
FIGURE 6.4. *The rank of the numerical solution for different tolerances ε.*

TABLE 6.1
Maximum number of columns of the basis matrix V_k along the iteration for the substepping approach and for the one step approximation using Algorithm 1 when an error tolerance ε is required.

ε	time stepping	single step iteration
10^{-2}	60	112
10^{-4}	76	147
10^{-6}	112	175
10^{-8}	160	203

than the error arising from the Krylov subspace approximation, we approximate the total error as

$$(6.2) \qquad \|X(Nh) - X_N\| \approx \sum_{\ell=1}^{N} \varepsilon_\ell,$$

where X_N denotes the numerical approximation after N steps. Figure 6.5 displays the error arising from the best 2-norm approximation after each step, i.e., the singular value σ_{41}, and the estimate (6.2). We see that the error at the end is not far from $900 \cdot \sigma_{41}^{(900)}$, the number of time steps times the largest rank cut.

6.2. Case $n = 5177$. Next we consider a finite element discretization with $n = 5177$. Figure 6.6 displays the convergence of Algorithm 1 and an a posteriori error estimate given by (5.7), when $T = 5$. For the scaling and squaring part we set the parameter $m = 10$.

We next carry out a time integration up to $T = 2000$ using $N = 1000$ time steps. We estimate the total error without access to a reference solution using the estimate (6.2). We use a Krylov subspace dimension $k = 32$ for the first time step and $k = 20$ for the rest of the steps and cut the rank to 40 after each step using the SVD. We see from Figure 6.7 that the a posteriori error estimate of Algorithm 1 is negligible compared to the error arising from the rank cut. We see that the estimate (6.2) is of the same order ($\approx$ ten times larger) as in the $n = 1357$-case.
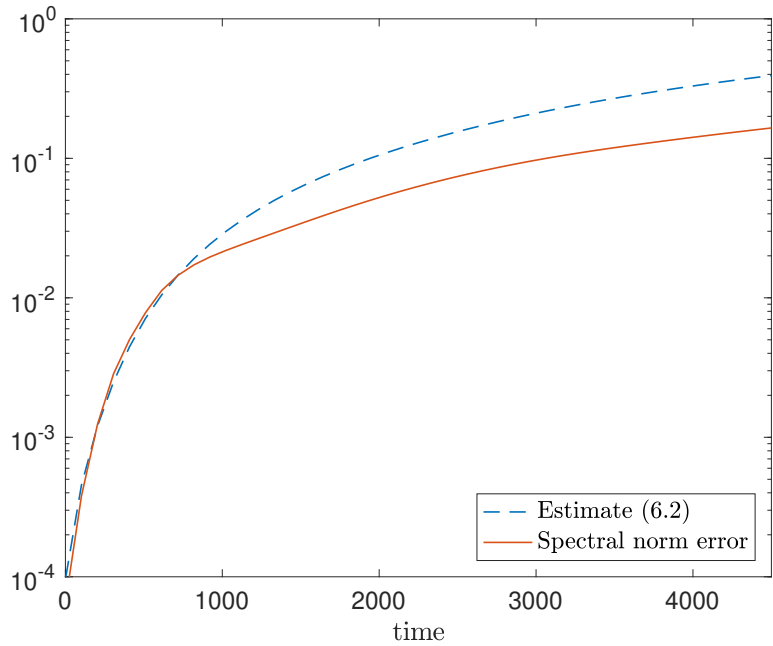
FIGURE 6.5. *The spectral norm error of the approximation and the estimate* (6.2) *for time integration up to* $T = 4500$. *The relative spectral norm error* $\|\widetilde{X}(T) - X(T)\|/\|X(T)\| \approx 4.7 \cdot 10^{-4}$ *at* $T = 4500$.
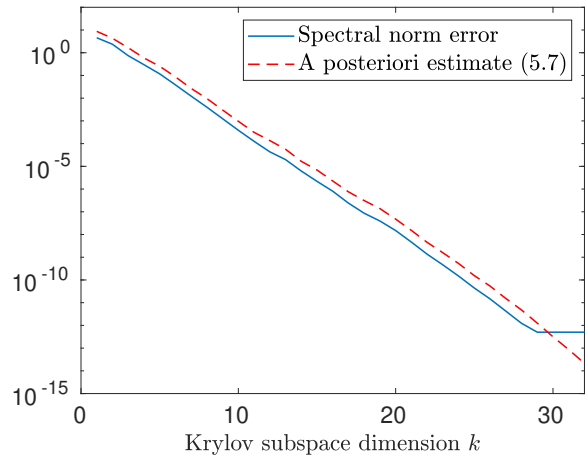

FIGURE 6.6. *Convergence of the approximation given by Algorithm 1 (for* $T = 5$*) and its a posteriori estimate* (5.7), *when* $n = 5177$.

7. Conclusions and Outlook. We have proposed a Krylov subspace approximation method for large-scale differential Riccati equations. We have shown that the method is structure preserving in the sense that it preserves two important properties of the exact flow, namely the property of positivity and under certain conditions also the property of monotonicity. We have also provided an a priori error analysis of the Krylov subspace approximation which
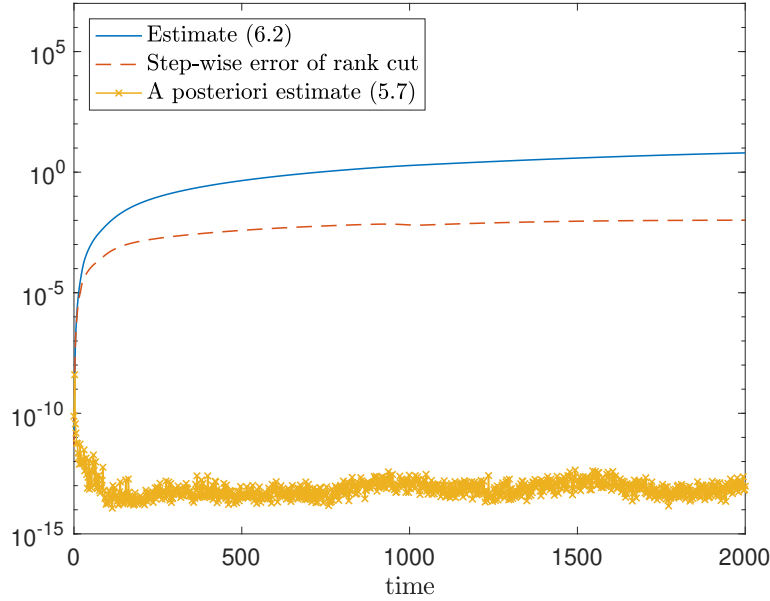
FIGURE 6.7. *The estimate* (6.2) *of the spectral norm error, size of the rank cut at each time step and the a posteriori estimate* (5.7) *at each time step when* $n = 5177$ *and a time integration up to* $T = 2000$.

shows superlinear convergence. This behavior was also verified in numerical experiments. In addition, an a posteriori error analysis was carried out, and the proposed estimate was shown to be accurate in numerical examples. In order to limit the memory consumption, in experiments we also considered limiting the rank of the numerical solution in multiple time stepping. To avoid excessively large approximation basis V_k, more studies of rational Krylov subspace methods are needed. Their benefits were illustrated in numerical experiments.

We would like to point out that the presented block Krylov subspace method can be extended to the non-symmetric differential Riccati equation. A possible extension could also be the non-autonomous case, i.e, the case in which the coefficient matrices Q, S, and A are time dependent. In this case an essential tool would be the so called Magnus expansion (see, e.g., [2]), which gives the fundamental solution of the linear system corresponding to the time-dependent coefficient matrix A.

Acknowledgments. The authors thank Valeria Simoncini for pointing out relevant literature related to the algebraic Riccati equation and Tony Stillfjord for several helpful comments on a draft of the paper. Moreover, the authors would like to thank the anonymous reviewers for their comments that greatly contributed to improving the paper.

Appendix A. Auxiliary Lemmas and the proof of Theorem 4.5.

We first state two auxiliary results needed for the proof of Theorem 4.5.

LEMMA A.1. *Let* $A \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{n \times \ell}$, *let* V_k *be a matrix with orthonormal columns such that* $\mathcal{K}_k(A, B) \subset \mathrm{R}(V_k)$, *and let* $H_k = V_k^T A V_k$. *Then, for all* $t, s > 0$ *it holds that*

$$\left\| \left(e^{tA} - V_k e^{tH_k} V_k^T \right) e^{sA} B \right\| \leq 4 \max \left\{ 1, e^{(t+s)\mu(A)} \right\} \frac{\|(t+s)A\|^k}{k!} \|B\|.$$

Proof. Using the polynomial approximation property of the Krylov approximation (see [30, Lemma 3.1]), we see that

$$
\begin{aligned}
V_k V_k^T e^{sA} B &= V_k V_k^T \sum_{\ell=0}^{k-1} \frac{(sA)^\ell}{\ell!} B + V_k V_k^T \sum_{\ell=k}^{\infty} \frac{(sA)^\ell}{\ell!} B \\
&= V_k \sum_{\ell=0}^{k-1} \frac{(sH_k)^\ell}{\ell!} V_k^T B + V_k V_k^T \sum_{\ell=k}^{\infty} \frac{(sA)^\ell}{\ell!} B \\
&= V_k e^{sH_k} V_k^T B - V_k \sum_{\ell=k}^{\infty} \frac{(sH_k)^\ell}{\ell!} V_k^T B + V_k V_k^T \sum_{\ell=k}^{\infty} \frac{(sA)^\ell}{\ell!} B.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\left(e^{tA} - V_k e^{tH_k} V_k^T \right) e^{sA} B \\
= e^{(t+s)A} B - V_k e^{tH_k} V_k^T (V_k V_k^T e^{sA} B) \\
= e^{(t+s)A} B - V_k e^{(t+s)H_k} V_k^T B + V_k e^{tH_k} \sum_{\ell=k}^{\infty} \frac{(sH_k)^\ell}{\ell!} V_k^T B \\
- V_k e^{tH_k} V_k^T \sum_{\ell=k}^{\infty} \frac{(sA)^\ell}{\ell!} B \\
= \sum_{\ell=k}^{\infty} \frac{((t+s)A)^\ell}{\ell!} B - V_k \sum_{\ell=k}^{\infty} \frac{((t+s)H_k)^\ell}{\ell!} V_k^T B \\
+ V_k e^{tH_k} \sum_{\ell=k}^{\infty} \frac{(sH_k)^\ell}{\ell!} V_k^T B - V_k e^{tH_k} V_k^T \sum_{\ell=k}^{\infty} \frac{(sA)^\ell}{\ell!} B.
\end{aligned}
$$

(A.1)

Using the bounds $\|e^{tA}\| \le e^{t\mu(A)}$, $\mu(H_k) \le \mu(A)$ and (see [14, Lemma A.2])

$$
\left\| \sum_{\ell=k}^{\infty} \frac{(tA)^\ell}{\ell!} \right\| \le \max\left\{ 1, e^{t\mu(A)} \right\} \frac{\|tA\|^k}{k!} \quad \text{for all} \quad t \ge 0
$$

for the four terms on the right-hand side of (A.1), the claim follows.   □

LEMMA A.2. *Let $X(s)$ be the solution of the Riccati differential equation* (2.1) *at time s, $0 \le s \le t$, and V_k be a matrix with orthonormal columns such that $\mathcal{K}_k(A, [Z\ C^T]) \subset \mathrm{R}(V_k)$. Denote $H_k = V_k^T A V_k$. Then, the following bound holds:*

$$
\begin{aligned}
\| \left(e^{(t-s)A} - V_k e^{(t-s)H_k} V_k^T \right) X(s) \| \\
\le 4\, c(s) \max\left\{ 1, e^{(t+s)\mu(A)} \right\} \|A\|^k \left(\frac{t^k}{k!} \|X_0\| + \frac{t^{k+1}}{(k+1)!} \|Q\| \right),
\end{aligned}
$$

where

$$
c(s) := 1 + s\|S\| \max_{w \in [0,s]} \|X(w)\| \, \varphi_1 \left(s\|S\| \max_{w \in [0,s]} \|X(w)\| \max\left\{ 1, e^{s\mu(A)} \right\} \right).
$$

Proof. Using the integral representation (2.4) for $X(s)$ we may write

$$
\left(e^{(t-s)A} - V_k e^{(t-s)H_k} V_k^T \right) X(s) = c_{1,k}(t,s) + c_{2,k}(t,s),
$$

where

$$c_{1,k}(t,s) = \left[\left(\mathrm{e}^{(t-s)A} - V_k \mathrm{e}^{(t-s)H_k} V_k^T \right) \mathrm{e}^{sA} Z \right] Z^T \mathrm{e}^{sA^T}$$

(A.2)
$$+ \int_0^s \left[\left(\mathrm{e}^{(t-s)A} - V_k \mathrm{e}^{(t-s)H_k} V_k^T \right) \mathrm{e}^{(s-u)A} C^T \right] C \mathrm{e}^{(s-u)A^T} \, \mathrm{d}u,$$

and

(A.3) $$c_{2,k}(t,s) = \left(\mathrm{e}^{(t-s)A} - V_k \mathrm{e}^{(t-s)H_k} V_k^T \right) \int_0^s \mathrm{e}^{(s-u)A} X(u) S X(u) \mathrm{e}^{(s-u)A^T} \, \mathrm{d}u.$$

By using Lemma A.1, we obtain for the expressions inside the square brackets on the right-hand side of (A.2) the bounds

$$\left\| \left[\left(\mathrm{e}^{(t-s)A} - V_k \mathrm{e}^{(t-s)H_k} V_k^T \right) \mathrm{e}^{sA} Z \right] Z^T \mathrm{e}^{sA^T} \right\|$$
$$\leq 4 \max \left\{ 1, \mathrm{e}^{(t+s)\mu(A)} \right\} \|X_0\| \|A\|^k \frac{t^k}{k!},$$

and

$$\left\| \int_0^s \left[\left(\mathrm{e}^{(t-s)A} - V_k \mathrm{e}^{(t-s)H_k} V_k^T \right) \mathrm{e}^{(s-u)A} C^T \right] C \mathrm{e}^{(s-u)A^T} \, \mathrm{d}u \right\|$$
$$\leq 4\|Q\| \int_0^s \frac{((t-u)\|A\|)^k}{k!} \max \left\{ 1, \mathrm{e}^{(t-u)\mu(A)} \right\} \max \left\{ 1, \mathrm{e}^{(s-u)\mu(A)} \right\} \, \mathrm{d}u$$
$$\leq 4\|Q\| \max \left\{ 1, \mathrm{e}^{(t+s)\mu(A)} \right\} \|A\|^k \frac{t^{k+1}}{(k+1)!}.$$

Thus,

(A.4) $$\|c_{1,k}(t,s)\| \leq 4 \max \left\{ 1, \mathrm{e}^{(t+s)\mu(A)} \right\} \|A\|^k \left(\frac{t^k}{k!} \|X_0\| + \frac{t^{k+1}}{(k+1)!} \|Q\| \right).$$

From (A.3) we see that

$$\|c_{2,k}(t,s)\| \leq \int_0^s \left\| \left(\mathrm{e}^{(t-s)A} - V_k \mathrm{e}^{(t-s)H_k} V_k^T \right) \mathrm{e}^{(s-u)A} X(u) \right\|$$

(A.5)
$$\times \|S\| \max_{w \in [0,s]} \|X(w)\| \mathrm{e}^{(s-u)\mu(A)} \, \mathrm{d}u.$$

Next we bound the first factor in the integrand of (A.5). We substitute the integral representation (2.4) for $X(u)$ to find that

$$
\begin{aligned}
& \left\| (e^{(t-s)A} - V_k e^{(t-s)H_k} V_k^T) e^{(s-u)A} X(u) \right\| \\
& \quad \leq \left\| \left[(e^{(t-s)A} - V_k e^{(t-s)H_k} V_k^T) e^{sA} Z \right] Z^T e^{uA^T} \right\| \\
& \qquad + \int_0^u \left\| \left[(e^{(t-s)A} - V_k e^{(t-s)H_k} V_k^T) e^{(s-w)A} C^T \right] C e^{(u-w)A^T} \right\| \, \mathrm{d}w \\
& \qquad + \int_0^u \left\| (e^{(t-s)A} - V_k e^{(t-s)H_k} V_k^T) e^{(s-w)A} X(w) \right\| \\
& \qquad\qquad \times \|S\| \max_{w \in [0,u]} \|X(w)\| \max\left\{ 1, e^{(u-w)\mu(A)} \right\} \, \mathrm{d}w.
\end{aligned}
$$
(A.6)

As above when bounding $\|c_{1,k}(t,s)\|$, we use Lemma A.1 for the expressions inside the square brackets on the right-hand side of (A.6), to get the inequality

$$
\begin{aligned}
& \left\| (e^{(t-s)A} - V_k e^{(t-s)H_k} V_k^T) e^{(s-u)A} X(u) \right\| \\
& \quad \leq 4\|A\|^k \max\left\{ 1, e^{(t+u)\mu(A)} \right\} \left(\frac{t^k}{k!} \|X_0\| + \frac{t^{k+1}}{(k+1)!} \|Q\| \right) \\
& \qquad + \|S\| \max_{w \in [0,u]} \|X(w)\| \max\left\{ 1, e^{u\mu(A)} \right\} \\
& \qquad\qquad \times \int_0^u \left\| (e^{(t-s)A} - V_k e^{(t-s)H_k} V_k^T) e^{(s-w)A} X(w) \right\| \, \mathrm{d}w.
\end{aligned}
$$
(A.7)

Applying Grönwall's lemma to (A.7), we find that

$$
\begin{aligned}
& \left\| (e^{(t-s)A} - V_k e^{(t-s)H_k} V_k^T) e^{(s-u)A} X(u) \right\| \\
& \quad \leq 4\|A\|^k \max\left\{ 1, e^{(t+u)\mu(A)} \right\} \left(\frac{t^k}{k!} \|X_0\| + \frac{t^{k+1}}{(k+1)!} \|Q\| \right) \\
& \qquad \times e^{u\|S\| \max_{w \in [0,u]} \|X(w)\| \max\{1, e^{u\mu(A)}\}}.
\end{aligned}
$$
(A.8)

Substituting (A.8) into (A.5), we get

$$
\begin{aligned}
& \|c_{2,k}(t,s)\| \\
& \quad \leq 4\|S\| \max_{w \in [0,s]} \|X(w)\| \|A\|^k \max\left\{ 1, e^{(t+s)\mu(A)} \right\} \\
& \qquad \times \left(\frac{t^k}{k!} \|X_0\| + \frac{t^{k+1}}{(k+1)!} \|Q\| \right) s\varphi_1\left(s\|S\| \max_{w \in [0,s]} \|X(w)\| \max\left\{ 1, e^{s\mu(A)} \right\} \right).
\end{aligned}
$$
(A.9)

The bounds (A.4) and (A.9) together show the claim. $\square$

Using Lemmas A.1 and A.2 we are now ready to prove Theorem 4.5.

Proof of Theorem 4.5. From the integral representation (2.4) for $X(t)$ and for the solution $Y_k(t)$ of the small-dimensional system (3.5), we see that

$$\text{(A.10)} \qquad X(t) - X_k(t) = F_{1,k}(t) + F_{2,k}(t),$$

where

$$F_{1,k}(t) := \mathrm{e}^{tA} X_0 \mathrm{e}^{tA^T} - V_k \mathrm{e}^{tH_k} V_k^T X_0 V_k \mathrm{e}^{tH_k^T} V_k^T$$
$$+ \int_0^t \left(\mathrm{e}^{(t-s)A} Q \mathrm{e}^{(t-s)A^T} - V_k \mathrm{e}^{(t-s)H_k} Q_k \mathrm{e}^{(t-s)H_k^T} V_k^T \right) \mathrm{d}s,$$

and

$$F_{2,k}(t) = \int_0^t \mathrm{e}^{(t-s)A} X(s) S X(s) \mathrm{e}^{(t-s)A^T} \, \mathrm{d}s$$

$$\text{(A.11)}$$

$$- \int_0^t V_k \mathrm{e}^{(t-s)H_k} V_k^T X_k(s) S X_k(s) V_k \mathrm{e}^{(t-s)H_k^T} V_k^T \, \mathrm{d}s.$$

Theorem 4.1 shows that $F_{1,k}(t)$ is bounded by

$$\text{(A.12)} \qquad \| F_{1,k}(t) \| \le 4 \max \left\{ 1, \mathrm{e}^{2t\mu(A)} \right\} \| A \|^k \left(\frac{t^k}{k!} \| X_0 \| + \frac{t^{k+1}}{(k+1)!} \| Q \| \right).$$

We add and subtract the term

$$\int_0^t \mathrm{e}^{(t-s)A} X(s) S X_k(s) V_k \mathrm{e}^{(t-s)H_k^T} V_k^T \, \mathrm{d}s$$

in (A.11) to obtain

$$F_{2,k}(t) = \int_0^t \mathrm{e}^{(t-s)A} X(s) S \, F_{3,k}(t,s)^T \, \mathrm{d}s + \int_0^t F_{3,k}(t,s) \, S \, X_k(s) V_k \mathrm{e}^{(t-s)H_k^T} V_k^T \, \mathrm{d}s,$$

where

$$F_{3,k}(t,s) = \mathrm{e}^{(t-s)A} X(s) - V_k \mathrm{e}^{(t-s)H_k} V_k^T X_k(s)$$
$$= \left(\mathrm{e}^{(t-s)A} - V_k \mathrm{e}^{(t-s)H_k} V_k^T \right) X(s) - V_k \mathrm{e}^{(t-s)H_k} V_k^T \left(X_k(s) - X(s) \right).$$

We see that

$$\text{(A.13)} \qquad \| F_{2,k}(t) \| \le 2 \| S \| \, \alpha(t) \int_0^t \max \left\{ 1, \mathrm{e}^{(t-s)\mu(A)} \right\}$$

$$\times \left(\| \left(\mathrm{e}^{(t-s)A} - V_k \mathrm{e}^{(t-s)H_k} V_k^T \right) X(s) \| + \| X(s) - X_k(s) \| \right) \, \mathrm{d}s,$$

where

$$\alpha(t) = \max \left\{ \max_{s \in [0,t]} \| X(s) \|, \max_{s \in [0,t]} \| X_k(s) \| \right\}.$$

The claim follows now from (A.10), (A.12), (A.13), Lemma A.1, Grönwall's lemma, Corollary 2.5, and Corollary 4.4, which form a sequence of substitutions.   ☐

REFERENCES

[1] H. ABOU-KANDIL, G. FREILING, V. IONESCU, AND G. JANK, *Matrix Riccati Equations*, Birkhäuser, Basel, 2003.

[2] P. BADER, S. BLANES, AND E. PONSODA, *Structure preserving integrators for solving (non-)linear quadratic optimal control problems with applications to describe the flight of a quadrotor*, J. Comput. Appl. Math., 262 (2014), pp. 223–233.

[3] B. BECKERMANN AND L. REICHEL, *Error estimates and evaluation of matrix functions via the Faber transform*, SIAM J. Numer. Anal., 47 (2009), pp. 3849–3883.

[4] P. BENNER AND H. MENA, *Numerical solution of the infinite-dimensional LQR problem and the associated Riccati differential equations*, J. Numer. Math., 26 (2018), pp. 1–20.

[5] ———, *Rosenbrock methods for solving Riccati differential equations*, IEEE Trans. Automat. Control, 58 (2013), pp. 2950–2956.

[6] P. BENNER AND J. SAAK, *A semi-discretized heat transfer model for optimal cooling of steel profiles*, in Dimension Reduction of Large-Scale Systems, P. Benner, D. C. Sorensen, V. Mehrmann, eds., Lecture Notes in Computational Science and Engineering 45, Springer, Berlin, 2005, pp. 353–356.

[7] M. A. BOTCHEV, V. GRIMM, AND M. HOCHBRUCK, *Residual, restarting, and Richardson iteration for the matrix exponential*, SIAM J. Sci. Comput., 35 (2013), pp. A1376–A1397.

[8] M. CORLESS AND A. FRAZHO, *Linear Systems and Control*, Marcel Dekker, New York, 2003.

[9] E. DAVISON AND M. MAKI, *The numerical solution of the matrix Riccati differential equation*, IEEE Trans. Automat. Control, 18 (1973), pp. 71–73.

[10] L. DIECI AND T. EIROLA, *Positive definiteness in the numerical solution of Riccati differential equations*, Numer. Math., 67 (1994), pp. 303–313.

[11] ———*Preserving monotonicity in the numerical solution of Riccati differential equations*, Numer. Math., 74 (1996), pp. 35–47.

[12] V. L. DRUSKIN AND L. A. KNIZHNERMAN, *Two polynomial methods for calculating functions of symmetric matrices*, Zh. Vychisl. Mat. i Mat. Fiz., 29 (1989), pp. 1763–1775.

[13] V. L. DRUSKIN, C. LIEBERMAN, AND M. ZASLAVSKY, *On adaptive choice of shifts in rational Krylov subspace reduction of evolutionary problems*, SIAM J. Sci. Comput., 32 (2010), pp. 2485–2496.

[14] E. GALLOPOULOS AND Y. SAAD, *Efficient solution of parabolic equations by Krylov approximation methods*, SIAM J. Sci. Statist. Comput., 13 (1992), pp. 1236–1264.

[15] Y. GULDOĞAN, M. HACHED, K. JBILOU, AND M. KURULAY, *Low-rank approximate solutions to large-scale differential matrix Riccati equations*, Appl. Math. (Warsaw), 45 (2018), pp. 233–254.

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

[17] S. GÜTTEL, *Rational Krylov approximation of matrix functions: numerical methods and optimal pole selection*, GAMM-Mitt., 36 (2013), pp. 8–31.

[18] N. J. HIGHAM, *Functions of Matrices*, SIAM, Philadelphia, 2008.

[19] ———, *The scaling and squaring method for the matrix exponential revisited*, SIAM J. Matrix Anal. Appl., 26 (2005), pp. 1179–1193.

[20] M. HOCHBRUCK AND C. LUBICH, *On Krylov subspace approximations to the matrix exponential operator*, SIAM J. Numer. Anal., 34 (1997), pp. 1911–1925.

[21] M. HOCHBRUCK AND A. OSTERMANN, *Exponential integrators*, Acta Numer., 19 (2010), pp. 209–286.

[22] A. ISERLES AND S. NØRSETT, *Order Stars*, Chapman & Hall, London, 1991.

[23] C. S. KENNEY AND R. B. LEIPNIK, *Numerical integration of the differential matrix Riccati equation*, IEEE Trans. Automat. Control, 30 (1985), pp. 962–970.

[24] L. KNIZHNERMAN AND V. SIMONCINI, *A new investigation of the extended Krylov subspace method for matrix function evaluations*, Numer. Linear Algebra Appl., 17 (2010), pp. 615–638.

[25] V. KUČERA, *A review of the matrix Riccati equation*, Kybernetika (Prague), 9 (1973), pp. 42–61.

[26] N. LANG, H. MENA, AND J. SAAK, *On the benefits of the LDL^T factorization for large-scale differential matrix equation solvers*, Linear Algebra Appl., 480 (2015), pp. 44–71.

[27] A. J. LAUB, *A Schur method for solving algebraic Riccati equations*, IEEE Trans. Automat. Control, 24 (1979), pp. 913–921.

[28] Y. LIN AND V. SIMONCINI, *Minimal residual methods for large scale Lyapunov equations*, Appl. Numer. Math., 72 (2013), pp. 52–71.

[29] L. LOPEZ AND V. SIMONCINI, *Preserving geometric properties of the exponential matrix by block Krylov subspace methods*, BIT, 46 (2006), pp. 813–830.

[30] Y. SAAD, *Analysis of some Krylov subspace approximations to the matrix exponential operator*, SIAM J. Numer. Anal., 29 (1992), pp. 209–228.

[31] ———, *Iterative Methods for Sparse Linear Systems*, PWS, Boston, 1996.

[32] ———, *Numerical solution of large Lyapunov equations*, in Signal processing, Scattering and Operator Theory, and Numerical Methods, M. A. Kaashoek, J. H. van Schuppen, and A. C. M. Ran, eds., Progress in Systems and Control Theory 5, Birkhäuser, Boston, 1990, pp. 503–511.

[33] V. SIMONCINI, *A new iterative method for solving large-scale Lyapunov matrix equations*, SIAM J. Sci. Comput., 29 (2007), pp. 1268–1288.

[34] ———, *Analysis of the rational Krylov subspace projection method for large-scale algebraic Riccati equations*, SIAM J. Matrix Anal. Appl., 37 (2016), pp. 1655–1674.

[35] T. STILLFJORD, *Adaptive high-order splitting schemes for large-scale differential Riccati equations*, Numer. Algorithms, 78 (2018), pp. 1129–1151.

[36] ———, *Low-rank second-order splitting of large-scale differential Riccati equations*, IEEE Trans Automat. Control, 60 (2015), pp. 2791–2796.

[37] L. TREFETHEN AND M. EMBREE, *Spectra and Pseudospectra*, Princeton University Press, Princeton, 2005.

[38] J. VAN DEN ESHOF AND M. HOCHBRUCK, *Preconditioning Lanczos approximations to the matrix exponential*, SIAM J. Sci. Comput., 27 (2006), pp. 1438–1457.