SINGULAR VALUE DECAY OF OPERATOR-VALUED DIFFERENTIAL LYAPUNOV AND RICCATI EQUATIONS

TONY STILLFJORD

Abstract. We consider operator-valued differential Lyapunov and Riccati equations, where the operators B and C may be relatively unbounded with respect to A (in the standard notation). In this setting, we prove that the singular values of the solutions decay fast under certain conditions. In fact, the decay is exponential in the negative square root if A generates an analytic semigroup and the range of C has finite dimension. This extends previous similar results for algebraic equations to the differential case. When the initial condition is zero, we also show that the singular values converge to zero as time goes to zero, with a certain rate that depends on the degree of unboundedness of C. A fast decay of the singular values corresponds to a low numerical rank, which is a critical feature in large-scale applications. The results reported here provide a theoretical foundation for the observation that, in practice, a low-rank factorization usually exists.

1. Introduction

We consider differential Lyapunov equations (DLEs) and differential Riccati equations (DREs) of the forms

$$\dot{P} = A^*P + PA + C^*C, \quad P(0) = G^*G, \tag{1}$$

and

$$\dot{P} = A^*P + PA + C^*C - PBB^*P, \quad P(0) = G^*G, \tag{2}$$

respectively. Such equations arise in many different areas, e.g. in optimal/robust control, optimal filtering, spectral factorizations, $\mathbf{H}_\infty$-control, differential games, etc. [1, 3, 17, 29].

A typical application for DREs is a linear quadratic regulator (LQR) problem, where one seeks to control the output $y = Cx$ given the state equation $\dot{x} = Ax + Bu$ by varying the input u. In the case of a finite time cost function,

$$J(u) = \int_0^T \|y(t)\|^2 + \|u(t)\|^2 \,\mathrm{d}t + \|Gy(T)\|^2,$$

it is well known that the optimal input function u^{opt} is given in state feedback form. In particular, $u^{\mathrm{opt}}(t) = -B^*P(T-t)x(t)$, where P is the solution to the DRE (2) [9, 20].

The solution to the DLE, on the other hand, yields the (time-limited) observability Gramian of the corresponding LQR system. It is used in applications such as model order reduction [5, 14] for determining which states x have negligible effect

Date: Received: date / Accepted: date.

2010 *Mathematics Subject Classification.* Primary 47A62; Secondary 47A11, 49N10.

Key words and phrases. Differential Riccati equations, differential Lyapunov equations, operator-valued, infinite-dimensional, singular value decay, low rank.

on the input-output relation $u \mapsto y$, and which can therefore safely be discarded from the system [18, 8].

In the continuous case, the equations (1), (2) are operator-valued. After a spatial discretization they become matrix-valued. Approximating their solutions by numerical computations is thus, if done naively, much more expensive than simply approximating, e.g., the corresponding vector-valued equation $\dot{x} = Ax$. A standard way to decrease the computational complexity is to utilize structural properties of the solutions. A commonly used such property is that of low numerical rank [21, 19, 36], i.e. a fast (often exponential) decay of the singular values. This allows us to approximate $P(t) \approx L(t)L(t)^*$ where $L(t)$ is of finite rank. In the matrix-valued setting, we would have $P(t) \in \mathbb{R}^{n\times n}$ and $L(t) \in \mathbb{R}^{n\times r}$ with $r \ll n$.

While there exist results on when such low numerical rank is to be expected for *algebraic* Lyapunov and Riccati equations (i.e. the stationary counterparts of (1) and (2)), see e.g. [2, 31, 4, 6, 28, 7, 15, 27], the differential case has so far been neglected in the literature.

The aim of this article is to remedy this situation and provide criteria on A, B and C that guarantee a certain decay of the singular values $\{\sigma_k\}_{k=1}^{\infty}$ of the solutions to (1) and (2). We consider the operator-valued case, with the standard assumption that A generates an analytic semigroup. In the LQR setting, this corresponds to the stable case, i.e. we exclude unstable but stabilizable problems. On the other hand, we allow relatively unbounded operators B and C, which means that we can treat various forms of boundary control and observation. In this setting, we follow the approach suggested in [27] for algebraic equations. There, a decay of the form $\sigma_k \leq C\mathrm{e}^{-\gamma\sqrt{k}}$ was shown, i.e. we can not expect exponential decay but only exponential in the square root. The main results of the present article demonstrates that this extends to the differential case, under similar assumptions. In the case that $G = 0$ (and hence $P(0) = 0$), our bounds additionally show that the singular values converge to 0 as $t \to 0$ with a rate $t^{1-2\alpha}$, where α is a measure of how unbounded the output operator C is.

An outline of the article is as follows: In Section 2 we specify the abstract framework, state the assumptions on the operators and recall some resulting properties of the solutions to (1) and (2). Then in Section 3 we use the concept of sinc quadrature to show that certain finite-rank operators approximate the integral $\int_0^t \left(C\mathrm{e}^{sA}\cdot, C\mathrm{e}^{sA}\cdot\right) \mathrm{d}s$ well. Since this is in fact the solution to (1) when $G = 0$, the main results for DLEs then follow quickly. We generalize these results to DREs in Section 4 by factorizing the system using output and input-output mappings. Finally, in Section 5, we perform a number of numerical experiments on discretized versions of the equations, which verify the theoretical statements.

2. Preliminaries

In the operator-valued case, (1) and (2) need to be interpreted in an appropriate sense. Here, we mainly follow [20] (see also [10]), and outline the ideas for the DRE (2) since all the results carry over to the DLE (1) by setting $B = 0$. Thus, let H, Y, U and Z be Hilbert spaces, and let the following operators be given: the (unbounded) state operator $A : \mathcal{D}(A) \subset H \to H$, the input operator $B : U \to \mathcal{D}(A^*)'$, the output operator $C : \mathcal{D}(A) \to Y$ and the final state penalization operator $G : H \to Z$. This corresponds to problems arising from the linear quadratic regulator setting.

By A^* we mean the adjoint of A with respect to the inner product on H, and $\mathcal{D}(A^*)'$ denotes the dual space of $\mathcal{D}(A)$, also with respect to the H-topology. With the proper interpretation (see e.g. [20]), it is a superset of H; in fact, the completion of H in the norm $\left\|A^{-1}\cdot\right\|_H$. Additionally, for general Hilbert spaces X and Y we use the notation $\mathcal{L}(X,Y)$ to denote the set of linear bounded operators from X to Y.

Remark 1. *In order that the notation conforms to the usual evolution equation setting, we have changed the direction of time so that $P(0) = G^*G$ is the given condition rather than $P(T) = G^*G$ as in [20]. The only effect of this is to change the signs of all the terms on the right-hand-side.*

Our main assumption is

Assumption 1. *The operator $A : \mathcal{D}(A) \subset H \to H$ is the generator of an exponentially stable analytic semigroup e^{tA} on H.*

This means that there exists a $\delta \in (0, \pi/2]$ such that $z \mapsto \mathrm{e}^{zA}$ is analytic on the sector $\Delta_\delta = \{z \in C \ ; \ z \neq 0, |\arg(z)| < \delta\}$. We note that A^* is the generator of $\mathrm{e}^{tA^*} = (\mathrm{e}^{tA})^*$ and that the fractional powers $(-A)^\gamma$ of A are well defined.

Further, we allow both B and C to be unbounded operators, but not *too* unbounded. In particular,

Assumption 2. *The operator $B : U \to \mathcal{D}(A^*)'$ is relatively bounded in the sense that there is a $\beta \in [0, 1)$ such that $(-A)^{-\beta}B \in \mathcal{L}(U, H)$.*

Assumption 3. *The operator $C : \mathcal{D}((-A)^\alpha) \to Y$ is relatively bounded in the sense that $C(-A)^{-\alpha} \in \mathcal{L}(H, Y)$ for $0 \leq \alpha < \min(1 - \beta, 1/2)$, with the parameter β from Assumption 2.*

Remark 2. *In the DLE case, we have $B = 0$. Assumption 2 is thus always satisfied and there is no extra restriction on α in Assumption 3 except $\alpha \in [0, 1/2)$.*

Under Assumptions 1 to 3, there exists a function P such that for all $x \in H$ and $\epsilon > 0$ we have $P(\cdot)x \in C([\epsilon, T], H)$, and $P(t)^* = P(t) \geq 0$ for $t \in (0, T]$. By this notation, we mean that $P(\cdot)x$ is continuous in the H-topology on $[\epsilon, T]$ and that $P(t)$ is a self-adjoint and a positive semi-definite operator. Moreover, $P(t)$ satisfies (2) in the following weak sense: Let $x, y \in \mathcal{D}((-A)^\epsilon)$ for $\epsilon > 0$. Then for all $0 < t < T$ we have that

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(P(t)x, y\right) = \left(P(t)x, Ay\right) + \left(P(t)Ax, y\right) + \left(C^*Cx, y\right) - \left(B^*P(t)x, B^*P(t)y\right), \quad (3)$$

where the inner products are all on H except the last one which is on U. In particular, the above equation holds for every $x, y \in \mathcal{D}(A)$.

When $\beta \in [0, 1/2)$, we may take $\epsilon = 0$. In fact, the function P is then actually a classical solution to (2), i.e. on $(0, T]$ the operator $P \mapsto A^*P + PA$ may be extended to an operator in $\mathcal{L}(H)$, $\dot{P}$ exists in $\mathcal{L}(H)$, Equation (2) is satisfied on $(0, T]$ and $P(0) = G^*G$. When $\beta \in [1/2, 1)$, we additionally have to require that the initial condition provides a certain amount of smoothing to draw the same conclusion:

Assumption 4. *There exists a $\theta \geq 0$ such that $\theta > 2\beta - 1$ if $\beta \geq 1/2$ and $\theta = 0$ otherwise, for which it holds that $(-A^*)^\theta G^*G \in \mathcal{L}(H)$.*

Under Assumption 4, the solution is still classical. In addition, we get uniqueness (in the class of self-adjoint functions $P(t) \in \mathcal{L}(H)$ such that $B^*P(t)x$ is continuous on $(0,T]$ with a singularity at 0 of size at most $t^{\theta-\beta}$). In addition, the weak solution to (3) solves the following integral equation for all $x, y \in H$ and vice versa:

$$\begin{aligned}(P(t)x, y) = \left(G\mathrm{e}^{tA}x, G\mathrm{e}^{tA}y\right) + \int_0^t \left(C\mathrm{e}^{sA}x, C\mathrm{e}^{sA}y\right) \mathrm{d}s \\ - \int_0^t \left(B^*P(s)\mathrm{e}^{sA}x, B^*P(s)\mathrm{e}^{sA}y\right) \mathrm{d}s\end{aligned} \tag{4}$$

This is equivalent to saying that for every $x \in H$ we have

$$P(t)x = \mathrm{e}^{tA^*}G^*G\mathrm{e}^{tA}x + \int_0^t \mathrm{e}^{sA^*}C^*C\mathrm{e}^{sA}x \,\mathrm{d}s - \int_0^t \mathrm{e}^{sA^*}P(s)BB^*P(s)\mathrm{e}^{sA}x \,\mathrm{d}s \tag{5}$$

We note that under Assumptions 1 and 3 it follows as in [27] that $C\mathrm{e}^{sA} \in \mathcal{L}(H,Y)$. In particular, for every $a \in [0,1)$ there exist positive constants M_a and ω such that $\left\|C\mathrm{e}^{zA}\right\|_{\mathcal{L}(H,Y)} \leq \frac{M_a}{|z|^\alpha}\mathrm{e}^{-\omega\Re(z)}$ for all $z \in \Delta_{a\delta}$. The constants M_a go to infinity as $a \to 1$, i.e. as we approach the limit of analyticity. However, by simply redefining δ as, e.g., $\delta/2$ we can always get a uniform estimate. In the following, we will therefore omit the dependence on a and write

$$\left\|C\mathrm{e}^{zA}\right\|_{\mathcal{L}(H,Y)} \leq \frac{M}{|z|^\alpha}\mathrm{e}^{-\omega\Re(z)}, \quad z \in \Delta_\delta, \tag{6}$$

for two positive constants M and ω. Since $\alpha < 1/2$, $|z|^{-2\alpha}$ is integrable at 0 and the first integral term in (5) is therefore well-defined. That the second integral term is well-defined under Assumptions 1 to 4 is less straightforward, due to the presence of $P(s)$ and the fact that β is allowed to take values in $[1/2, 1)$. We refer to [20, Chapter 1].

3. Lyapunov equations

Let us first consider the Lyapunov case (1). Restricting (4) by setting $B = 0$ shows that

$$(P(t)x, y) = \left(G\mathrm{e}^{tA}x, G\mathrm{e}^{tA}y\right) + \int_0^t \left(C\mathrm{e}^{sA}x, C\mathrm{e}^{sA}y\right) \mathrm{d}s, \tag{7}$$

which provides a closed-form expression for the solution P. For $x, y \in \mathcal{D}(A)$ we denote the integrand by F;

$$F(z) = \left(C\mathrm{e}^{zA}x, C\mathrm{e}^{zA}y\right), \tag{8}$$

and note that in fact $F : \Delta_\delta \to \mathbb{C}$. By (6), for all $x, y \in \mathcal{D}(A)$ we have the bound

$$|F(z)| < \frac{M^2}{|z|^{2\alpha}}\mathrm{e}^{-2\omega\Re(z)} \|x\| \|y\| \tag{9}$$

Our aim is now to approximate the integral $\int_0^t F(s) \,\mathrm{d}s$ by sinc quadrature, which converges exponentially in the number of quadrature nodes. The basic idea is to map the interval $(0,t)$ onto the real line, apply the trapezoidal rule, use decay properties of F at $\pm\infty$ and then transform back. The proof uses complex analysis and thus requires us to consider $(0,t)$ as a subset of a domain in $\mathbb{C}$ rather than a

real interval. In our case, the appropriate mapping is $\phi\colon \mathbb{C} \to \mathbb{C}$, $\phi(z) = \frac{z}{t-z}$, with inverse $\psi\colon \mathbb{C} \to \mathbb{C}$, $\psi(w) = \frac{te^w}{e^w+1}$. The function ϕ maps the eye-shaped domain

$$D_E^d(t) = \{z \in \mathbb{C}\ ;\ |\arg\left(\frac{z}{t-z}\right)| < d\},$$

where $0 < d < \pi/2$, onto the infinite strip

$$D_S^d(t) = \{w \in \mathbb{C}\ ;\ |\Im w| < d\}.$$

Here, of course, $D_E^d(t) \supset [0,t]$. The following result is due to Lund and Bowers [23], inspired by [34]. Here, as well as throughout the rest of the paper, we use the letter C to denote a generic constant in addition to denoting the output operator. The context makes it clear which interpretation is intended, and no confusion should arise.

Theorem 1 ([23, Theorem 3.8])**.** *Let f be an analytic function on $D_E^d(t)$ that for some $r \in (0,1)$ satisfies the condition*

$$\int_{\Psi(u+L)} |f(z)\,\mathrm{d}z| = \mathcal{O}(|u|^r), \quad u \to \pm\infty, \tag{10}$$

where $L = \{iv\ ;\ |v| \le d\}$. Further assume that

$$B(f) := \lim_{\gamma \to \partial D_E^d(t)} \int_\gamma |f(z)\,\mathrm{d}z| < \infty, \tag{11}$$

where γ denotes any closed simple contour in $D_E^d(t)$, and that there are positive constants C, ρ and μ such that

$$\left|\frac{f(z)}{\phi'(z)}\right| \le C \begin{cases} e^{-\rho|\phi(z)|} & \forall z \in \psi\big((-\infty,0)\big) \\ e^{-\mu|\phi(z)|} & \forall z \in \psi\big([0,-\infty)\big) \end{cases}. \tag{12}$$

Choose

$$n = \left\lceil \frac{\rho}{\mu} m + 1 \right\rceil, \quad h = \left(\frac{2\pi d}{\rho m}\right)^{1/2},$$

with m a nonnegative integer large enough that $h \le \frac{2\pi d}{\log 2}$, and define the quadrature nodes z_k and weights w_k by

$$z_k = \psi(kh) = \frac{te^{kh}}{e^{kh}+1}, \quad w_k = \Big(\phi'(z_k)\Big)^{-1} = \frac{te^{kh}}{(e^{kh}+1)^2}.$$

Then it holds that

$$\left| \int_0^t f(z)\,\mathrm{d}z - h \sum_{k=-m}^{n} w_k f(z_k) \right| \le \left(\frac{C}{\rho} + \frac{C}{\mu} + 2B(f)\right) e^{-(2\pi\rho d m)^{1/2}}.$$

Specifying this theorem to the function F given in (8) leads to

Theorem 2. *Let Assumptions 1 and 3 be satisfied, and let h, n, z_k and w_k be chosen as in Theorem 1. Then there is a positive constant C, independent of t, x and y, but dependent on α, such that*

$$\left| \int_0^t F(z)\,\mathrm{d}z - h \sum_{k=-m}^{n} w_k F(z_k) \right| \le C t^{1-2\alpha} e^{-(2\pi(1-2\alpha)\delta m)^{1/2}} \|x\|\,\|y\|\,.$$

Proof. We verify the conditions of Theorem 1. Since the domain $D_E^\delta(t)$ is a subset of the cone $\{w \in \mathbb{C}\ ;\ |\arg w| \leq \delta\}$for any $t > 0$, the function F is clearly analytic on $D_E^\delta(t)$. Suppose that $z = \psi(u + iv)$ where $|v| \leq \delta$. Then

$$\left|\frac{\mathrm{d}z}{\mathrm{d}v}\right| = \frac{t\mathrm{e}^u}{|\mathrm{e}^u\mathrm{e}^{iv} + 1|^2} \leq t\min(\mathrm{e}^u, \mathrm{e}^{-u}) \leq t,$$

since $\delta < \pi/2$ means that $|\mathrm{e}^u\mathrm{e}^{iv} + 1| \geq \max(1, \mathrm{e}^u)$. Hence

$$\begin{aligned}\int_{\Psi(u+L)} |F(z)\,\mathrm{d}z| &\leq \int_{-\delta}^{\delta} \left|F\left(\frac{t\mathrm{e}^u\mathrm{e}^{iv}}{\mathrm{e}^u\mathrm{e}^{iv} + 1}\right)\right| t\ \mathrm{d}v \\ &\leq Ct\int_{-\delta}^{\delta} \left|\frac{t\mathrm{e}^u\mathrm{e}^{iv}}{\mathrm{e}^u\mathrm{e}^{iv} + 1}\right|^{-2\alpha} \mathrm{d}v \\ &\leq 2C\pi t^{1-2\alpha},\end{aligned}$$

where we have used (9) as well as the coarse estimate $\mathrm{e}^{-2\omega\Re(z)} \leq 1$ in the second step and the inequality $|\mathrm{e}^u\mathrm{e}^{iv} + 1| \leq \mathrm{e}^u + 1 \leq 2\mathrm{e}^u$ in the third step. As this bound is independent of u and $1 - 2\alpha > 0$ due to Assumption 3, the first condition (10) is satisfied.

To check the second condition, we make a change of variables $w = \eta(z) = \frac{z}{t-z}$. It is easily seen that η maps the boundary of $D_E^\delta(t)$ onto the rays $\{r\mathrm{e}^{\pm i\delta}\ ;\ r \geq 0\}$, that the inverse is given by $z = \eta^{-1}(w) = \frac{tw}{1+w}$ and that the derivative of the inverse is given by $w \mapsto \frac{t}{(1+w)^2}$. Denoting the top and bottom parts of $\partial D_E^\delta(t)$ by ∂D_+ and ∂D_-, respectively, we thus have $B(F) = \int_{\partial D_+} |F(z)\,\mathrm{d}z| + \int_{\partial D_-} |F(z)\,\mathrm{d}z|$ where

$$\begin{aligned}\int_{\partial D_\pm} |F(z)\,\mathrm{d}z| &= \int_0^\infty \left|F\left(\frac{tr\mathrm{e}^{\pm i\delta}}{1 + r\mathrm{e}^{\pm i\delta}}\right)\right| t\big|1 + r\mathrm{e}^{\pm i\delta}\big|^{-2}\,\mathrm{d}r \\ &\leq C\int_0^\infty \left|\frac{tr\mathrm{e}^{\pm i\delta}}{1 + r\mathrm{e}^{\pm i\delta}}\right|^{-2\alpha} t\big|1 + r\mathrm{e}^{\pm i\delta}\big|^{-2}\,\mathrm{d}r,\end{aligned}$$

again using (9) and bounding the exponential term by 1. Since $|1 + r\mathrm{e}^{\pm i\delta}| \geq \max(1, r)$ we get

$$\int_{\partial D_\pm} |F(z)\,\mathrm{d}z| \leq t^{1-2\alpha}\left(\int_0^1 r^{-2\alpha}\,\mathrm{d}r + \int_1^\infty r^{-2}\,\mathrm{d}r\right),$$

so that, in conclusion,

$$B(F) \leq 2t^{1-2\alpha}\Big(\frac{1}{1-2\alpha} + 1\Big).$$

Finally, we check condition (12). A simple computation shows that $\phi'(z) = \frac{t}{z(t-z)}$. Clearly, $\psi\big((-\infty, 0)\big) = (0, t/2) =: \Gamma_1$ and $\psi\big([0, \infty)\big) = [t/2, t) =: \Gamma_2$, which means that on these intervals we have

$$\mathrm{e}^{-\rho|\phi(z)|} = z^\rho(t - z)^{-\rho} \quad \text{and} \quad \mathrm{e}^{-\mu|\phi(z)|} = z^{-\mu}(t - z)^\mu.$$

On Γ_1, $|t - z| \leq t$, so by (9) we get

$$\begin{aligned}\left|\frac{F(z)}{\phi'(z)}\right| &\leq C|z|^{-2\alpha}\mathrm{e}^{-2\omega\Re(z)}|z||t - z|t^{-1} \leq C|z|^{1-2\alpha}t^{-1}|t - z|^{2\alpha-1}|t - z|^{2-2\alpha} \\ &\leq Ct^{1-2\alpha}|z|^{1-2\alpha}|t - z|^{2\alpha-1},\end{aligned}$$

i.e. the desired bound holds with $\rho = 1 - 2\alpha$ and constant $Ct^{1-2\alpha}$, where C is independent of t. On Γ_2, $|z| \leq t$, and we similarly get

$$\begin{aligned}\left|\frac{F(z)}{\phi'(z)}\right| &\leq C|z|^{1-2\alpha}|t-z|t^{-1} \leq C|z|^{-1}|t-z||z|^{2-2\alpha}t^{-1} \\ &\leq Ct^{1-2\alpha}|z|^{-1}|t-z|,\end{aligned}$$

i.e. the desired bound holds with $\mu = 1$ and constant $Ct^{1-2\alpha}$, where C is again independent of t. □

Remark 3. *In the current approach, the factor $t^{1-2\alpha}$ is desired when t is small, but also means that the bound deteriorates when $t \to \infty$. We can use the factor $\mathrm{e}^{-2\omega\Re(z)}$, which we previously estimated by 1, to compensate this. However, this only works on (e.g.) the interval $[t/2, t)$: we can not bound $\mathrm{e}^{-2\omega\Re(z)}$ uniformly on $(0, t/2)$ by e^{-Ct} for any positive C.*

If $t \in [0, T]$ where T is very large, we might instead utilize [35, Example 4.2.10] *for the infinite interval. In this case, the new choice of mapping ϕ means that it is straightforward to acquire the necessary decay condition by using the $\mathrm{e}^{-2\omega\Re(z)}$ term. The same idea, but in a different formulation, is presented in* [23] *and used in* [27] *with the function F to show the exponential square-root decay*

$$\left|\int_0^\infty F(z)\,\mathrm{d}z - h\sum_{k=-m}^{n} F(\mathrm{e}^{kh})\mathrm{e}^{kh}\right| \leq C\mathrm{e}^{-\sqrt{2\pi\delta\alpha m}}.$$

By (9) we have

$$\left|\int_0^T F(z)\,\mathrm{d}z - \int_0^\infty F(z)\,\mathrm{d}z\right| \leq \frac{CT^{-2\alpha}\mathrm{e}^{-2\omega T}}{2\omega},$$

and we thus get exponential square-root decay except for a small constant term, if T is large.

We denote the singular values of P by $\sigma_k(P)$ and order them in decreasing order. Let us first consider the case when $G = 0$.

Theorem 3. *Let Assumptions 1 and 3 be satisfied, with the output space Y having finite dimension. Further assume that $G = 0$. Then the singular values of the solution P to the DLE (7) satisfy*

$$\sigma_k(P(t)) \leq Ct^{1-2\alpha}\mathrm{e}^{-\eta\sqrt{k+1-\dim Y}},$$

for $k \geq \max(1, \dim Y - 1)$, where C and η are positive constants independent of t but dependent on α.

After our preliminary work, the proof follows almost exactly as in [27]:

Proof. We have

$$(P(t)x, y) = \int_0^t F(z)\,\mathrm{d}z.$$

Now define n, z_k and w_k as in Theorem 2 and define the approximation P_m by

$$P_m = h\sum_{k=-m}^{n} w_k \mathrm{e}^{z_k A^*} C^* C \mathrm{e}^{z_k A}.$$

Since $P(t)$ and $P_m(t)$ are both self-adjoint operators and $\mathcal{D}(A)$ is dense in H, by Theorem 2 we then get

$$\begin{aligned}\|P(t)-P_m(t)\| &= \sup_{\substack{z\in\mathcal{D}(A)\\ \|z\|=1}} \big|\big(\big(P(t)-P_m(t)\big)z,z\big)\big| \\ &\leq Ct^{1-2\alpha}\mathrm{e}^{-(2\pi(1-2\alpha)dm)^{1/2}}.\end{aligned}$$

Now let

$$k_m = (2m+1)\dim Y.$$

Since $n \leq m$, the rank of $P_m(t)$ is at most k_m and we immediately see that

$$\sigma_{k_m+1}(P(t)) \leq Ct^{1-2\alpha}\mathrm{e}^{-(2\pi(1-2\alpha)dm)^{1/2}}.$$

As the singular values are decreasing, we may rewrite this [1] as

$$\sigma_j \leq \tilde{C}t^{1-2\alpha}\mathrm{e}^{-\eta\sqrt{j+1-\dim Y}},$$

for $j \geq \dim Y - 1$, with the modified constants $\tilde{C} = C\mathrm{e}^{(2\pi(1-2\alpha)d)^{1/2}(\sqrt{2}-1)}$ and $\eta = (2\pi(1-2\alpha)d)^{1/2}/\sqrt{2\dim Y}$. □

Remark 4. *We could clearly also estimate $k_m \leq m \cdot 3\dim Y$ and thereby get $\sigma_j \leq \tilde{C}t^{1-2\alpha}\mathrm{e}^{-\hat{\eta}\sqrt{j}}$ for $j \geq 1$ (with the same $\tilde{C}$). However, the estimated decay rate is then instead*

$$\hat{\eta} = (2\pi(1-2\alpha)d)^{1/2}/\sqrt{3\dim Y}$$

which is worse than the bound given in Theorem 3.

A non-zero operator G makes the situation more delicate. If G is a finite-rank operator, then the above result is essentially just shifted by $\operatorname{rank}(G)$. For consistency, we formulate this in terms of the output space Z:

Theorem 4. *Let Assumptions 1, 3 and 4 be satisfied, with the output spaces Y and Z both having finite dimension. Then the singular values of the solution P to the DLE (7) satisfy*

$$\sigma_k(P(t)) \leq Ct^{1-2\alpha}\mathrm{e}^{-\eta\sqrt{k+1-\dim Y-\dim Z}},$$

for $k \geq \max(1, \dim Y + \dim Z - 1)$, where C and η are positive constants independent of t but dependent on α.

Proof. This follows by the same procedure as in the proof of Theorem 3 after changing the definition of P_m to

$$P_m = \mathrm{e}^{tA^*}G^*G\mathrm{e}^{tA} + h\sum_{k=-m}^{n} w_k \mathrm{e}^{z_k A^*}C^*C\mathrm{e}^{z_k A}.$$

In this case, $k_m = \dim Z + (2m+1)\dim Y$.

Alternatively, we may use the well-known Weyl's inequality (also known as the Ky Fan inequality): Let F_1 and F_2 be two compact operators on H with singular values $\{\sigma_k^1\}_{k=1}^\infty$ and $\{\sigma_k^2\}_{k=1}^\infty$, respectively. Denote the singular values of $F_1 + F_2$ by $\{\sigma_k\}_{k=1}^\infty$. Then $\sigma_{j+k-1} \leq \sigma_j^1 + \sigma_k^2$ for all positive integers j and k [13].

[1] Let $k = a + bm$ with $a = \dim Y$ and $b = 2\dim Y$. For $j = k, k+1, \dots, k+b$ we have $\sigma_j \leq \sigma_k \leq C\mathrm{e}^{-\tilde{\gamma}\sqrt{k-a}} \leq C\mathrm{e}^{-\tilde{\gamma}\sqrt{j-a}}\mathrm{e}^{-\tilde{\gamma}\left(\sqrt{k-a}-\sqrt{j-a}\right)}$, with $\tilde{\gamma} = \gamma/\sqrt{b}$. Now, $\sqrt{k-a}-\sqrt{j-a} \geq \sqrt{k-a}-\sqrt{k+b-a} = \sqrt{bm}-\sqrt{b(m+1)}$. The latter function attains its minimum at $m=1$, so we get $\tilde{\gamma}(\sqrt{k-a}-\sqrt{j-a}) \geq 1-\sqrt{2}$. Hence, $\sigma_j \leq C\mathrm{e}^{\gamma(\sqrt{2}-1)}\mathrm{e}^{-\frac{\gamma}{\sqrt{b}}\sqrt{j-a}}$.

If $\dim Z < \infty$, then G and $\mathrm{e}^{tA^*}G^*G\mathrm{e}^{tA}$ are both compact operators whose singular values are zero except for the first $\dim Z$ ones. The operator $\int_0^t \mathrm{e}^{sA^*}C^*C\mathrm{e}^{sA}\,\mathrm{d}s$ is also compact, since it is the limit of a sequence of finite-rank operators (see the first part of the proof for Theorem 3). Hence Weyl's inequality applies, which shifts the start of the exponential decay by $\dim Z$. □

Finally, we consider the case where G is a general operator. To handle the term $\mathrm{e}^{tA^*}G^*G\mathrm{e}^{tA}$ we then have to impose stricter requirements on the semigroup e^{tA} and, by extension, its generator A.

Theorem 5. *Let Assumptions 1, 3 and 4 be satisfied, with the output space Y having finite dimension and $\dim Z = \infty$. Additionally, assume that the singular values of the solution operator e^{tA} decay exponentially in the square root; $\sigma_k(\mathrm{e}^{tA}) \leq \tilde{C}\mathrm{e}^{-\tilde{\eta}\sqrt{k}}$. Then the singular values of the solution P to the DLE (7) satisfy*

$$\sigma_k(P(t)) \leq C\max(1, t^{1-2\alpha})\mathrm{e}^{-\eta\sqrt{k+2-\dim Y}},$$

for $k \geq \max(1, \dim Y - 2)$, where C and η are positive constants independent of t but dependent on α.

Proof. The extra assumption on e^{tA} in particular implies that e^{tA} is compact, and since G^*G is a bounded operator also $\mathrm{e}^{tA^*}G^*G\mathrm{e}^{tA}$ is compact. We may therefore apply Weyl's inequality, as in the proof of Theorem 4. By Theorem 3 we directly get that

$$\begin{aligned}\sigma_{2k-\dim Y}(P(t)) &= \sigma_{k+(k+1-\dim Y)-1}(P(t)) \\ &\leq Ct^{1-2\alpha}\mathrm{e}^{-\eta\sqrt{k+1-\dim Y}} + \tilde{C}\mathrm{e}^{-\tilde{\eta}\sqrt{k+1-\dim Y}} \\ &\leq 2\max(Ct^{1-2\alpha}, \tilde{C})\mathrm{e}^{-\min(\eta,\tilde{\eta})\sqrt{k+1-\dim Y}}\end{aligned}$$

and thus

$$\sigma_j(P(t)) \leq 2\max(Ct^{1-2\alpha}, \tilde{C})\mathrm{e}^{-\frac{1}{2}\min(\eta,\tilde{\eta})\sqrt{j+2-\dim Y}}$$

for all $j \geq \max(1, \dim Y - 2)$. □

Remark 5. *When A is diagonalizable, the new assumption on e^{tA} obviously means that the eigenvalues of A should go to $-\infty$ like the negative square root. This assumption is satisfied in many concrete applications. As an example, the Laplacian on $\Omega \in \mathbb{R}^d$ with Dirichlet or Neumann boundary conditions has eigenvalues $\lambda_k(A)$ that decrease as $\lambda_k(A) = \mathcal{O}(-k^2/d)$ by Weyl's law, see e.g.* [12, Chapter VI]. *Hence the assumption is satisfied for such problems of up to dimension 2, while the decay for a 3D problem could drop to $\mathcal{O}\big(\mathrm{e}^{-\eta k^{1/3}}\big)$.*

4. Riccati equations

As in [27], we may extend the Lyapunov results to the Riccati case by using a factorization into output and input-output maps. For this, we will employ the framework of well-posed systems advocated by Salamon [30] and Staffans [32], see also [25, 37]. This restricts the class of problems compared to the setting of Lasiecka and Triggiani that we have used so far, in the sense that we need to make the state space larger to be able to consider problems with $\beta > 1/2$. We refer to [20, Section 3.3.3]. As an example, take the Dirichlet boundary control problem given

in [20, Section 3.3.1]. This is naturally[2] posed in $H = L^2(\Omega)$ with $\beta = 3/4 + \epsilon$ ($\epsilon > 0$ a small parameter) and $\alpha = 0$, but it is *not* well-posed on H, because there exist input functions $u \in L^2([0,T], U)$ which lead to outputs y that are only in $L^2([0,T], L^2(\Omega))$ and not in $C([0,T], Y)$. However, instead formulating the problem on, e.g., $H = H^{-1+2\epsilon}(\Omega)$ means that B becomes less unbounded and C more so: we get instead $\beta = 1/4 + 2\epsilon$ and $\alpha = 1/2 - \epsilon$. A better compromise might be $H = H^{-1/2-4\epsilon}(\Omega)$, which gives $\beta = 1/2 - \epsilon$ and $\alpha = 1/4 + 2\epsilon$. This is essentially a change of pivot space, cf. [33]. In either of these cases, we lose some (spatial) regularity of the solution to the DRE, but in return we may speak about output and input-output mappings.

As in Section 3 we first consider the case of a zero initial condition, then extend this to the finite-rank case and finally to the case of a general G but with extra requirements on A.

Theorem 6. *Let Assumptions 1 to 4 be satisfied, with the output spaces Y and Z having finite dimension. Then if $G = 0$, the singular values of the solution P to the DRE (4) satisfy*

$$\sigma_k(P(t)) \leq Ct^{1-2\alpha}\mathrm{e}^{-\eta\sqrt{k+1-\dim Y}},$$

for $k \geq \max(1, \dim Y - 1)$. If $G \neq 0$ but $\dim Z < \infty$ we instead get

$$\sigma_k(P(t)) \leq Ct^{1-2\alpha}\mathrm{e}^{-\eta\sqrt{k+1-\dim Y-\dim Z}},$$

for $k \geq \max(1, \dim Y + \dim Z - 1)$. Finally, if $\dim Z = \infty$ but $\sigma_k(\mathrm{e}^{tA}) \leq \tilde{C}\mathrm{e}^{-\tilde{\eta}\sqrt{k}}$, then

$$\sigma_{k(P(t))} \leq C\max(1, t^{1-2\alpha})\mathrm{e}^{-\eta\sqrt{k+2-\dim Y}},$$

for $k \geq \max(1, \dim Y - 2)$. In all the cases above, C and η are positive constants independent of t but dependent on α. Further, if $\beta > 1/2$, the results are with respect to an appropriately chosen pivot space $(-A)^{-\gamma}H$ rather than H.

Remark 6. *As in Remark 4, we can shift the decay to start at $k = 1$ at the expense of a lower decay rate η. This is also apparent from this given formulas. In the first case, for example, simply factoring out $\sqrt{k}$ yields the new decay rate $\eta\sqrt{\frac{\dim Y-2}{\dim Y-1}}$ when $\dim Y > 2$.*

Proof. If $\beta > 1/2$, we perform a change of pivot space from H to $(-A)^{-\gamma}H$ with an appropriate $\gamma > 0$, as discussed in the beginning of Section 4. We still denote this new space by H. By [32, Theorem 5.7.3], the system is then well-posed and regular in H. That is, the operators $\mathcal{A}(t)$, $\mathcal{B}(t)$, $\mathcal{C}(t)$ and $\mathcal{D}(t)$ defined by

$$\mathcal{A}(t)x_0 = \mathrm{e}^{tA}x_0, \qquad \mathcal{B}(t)u = \int_0^t \mathrm{e}^{(t-s)A}Bu(s)\,\mathrm{d}s,$$
$$\mathcal{C}(t)x_0 = C\mathrm{e}^{\cdot A}x_0 \qquad \mathcal{D}(t)u = \int_0^{\cdot} C\mathrm{e}^{(\cdot-s)A}Bu(s)\,\mathrm{d}s$$

satisfy $\mathcal{A}(t) \in \mathcal{L}(H,H)$, $\mathcal{B}(t) \in \mathcal{L}(L^2([0,t],U),H)$, $\mathcal{C}(t) \in \mathcal{L}(H, L^2([0,t],Y))$ and $\mathcal{D}(t) \in \mathcal{L}(L^2([0,t],U), L^2([0,t],Y))$. Here, $\mathcal{B}$ is known as the input map, $\mathcal{C}$ is the output map and $\mathcal{D}$ is the input-output map.

[2] In the sense that the Laplacian is more naturally seen as an unbounded operator on $L^2(\Omega)$, rather than on, say, $H^{-1}(\Omega)$.

When $G = 0$ we can directly apply the result of Salamon [30, Theorem 5.1], which (in our notation) states that

$$P(t) = \mathcal{C}(t)^* \big(\mathcal{I} + \mathcal{D}(t)\mathcal{D}(t)^*\big)^{-1} \mathcal{C}(t).$$

Here, $\mathcal{I}$ denotes the identity operator on $L^2([0,t], Y)$, and the inverse of $\mathcal{I} + \mathcal{D}(t)\mathcal{D}(t)^*$ exists as a bounded self-adjoint operator by the Lax-Milgram lemma. A straightforward calculation shows that $\mathcal{C}(t)^*$ is given by $\mathcal{C}(t)^* u = \int_0^t \mathrm{e}^{sA^*} C^* u(s)\,\mathrm{d}s$, and we thus have

$$\mathcal{C}(t)^*\mathcal{C}(t)x_0 = \int_0^t \mathrm{e}^{sA^*} C^* C \mathrm{e}^{sA} x_0 \,\mathrm{d}s.$$

Thus, in fact, for $x, y \in \mathcal{D}(A)$ we have $(\mathcal{C}(s)^*\mathcal{C}(s)x, y) = F(s)$ with F defined by (8). Hence the singular values of $\mathcal{C}(t)^*\mathcal{C}(t)$ decay exponentially in the square root, by exactly the same reasoning as in the proof of Theorem 3. Multiplying $\mathcal{C}(t)^*\mathcal{C}(t)$ by the bounded operator $(\mathcal{I} + \mathcal{D}\mathcal{D}^*)^{-1}$ only scales the singular values by the factor $\big\|(\mathcal{I} + \mathcal{D}\mathcal{D}^*)^{-1}\big\|$, so we have thus proven the first assertion.

The argument in [30, Theorem 5.1] may be extended also to the more general case that $G \neq 0$. We instead get

$$\begin{aligned} P(t) &= \mathcal{C}_G^*\mathcal{C}_G + \mathcal{C}(t)^*\mathcal{C}(t) \\ &\quad - \big(\mathcal{C}_G^*\mathcal{D}_G + \mathcal{C}(t)^*\mathcal{D}(t)\big)\big(\mathcal{I} + \mathcal{D}(t)\mathcal{D}(t)^* + \mathcal{D}_G\mathcal{D}_G^*\big)^{-1}\big(\mathcal{D}_G^*\mathcal{C}_G + \mathcal{D}(t)^*\mathcal{C}(t)\big), \end{aligned}$$

where

$$\mathcal{C}_G x_0 = G\mathrm{e}^{TA}x_0 \quad \text{and} \quad \mathcal{D}_G u = G \lim_{t\to T} \int_0^t \mathrm{e}^{(t-s)A} B u(s)\,\mathrm{d}s$$

are the "final-state" versions of the $\mathcal{C}$ and $\mathcal{D}$ operators. Recall that the problem is stated on $t \in [0,T]$. For any such t, we define the product space $X_t = L^2([0,t], Y) \times L^2([0,t], U) \times Z$ with the induced topology

$$\left\| \begin{bmatrix} y \\ u \\ z \end{bmatrix} \right\|_{X_t} = \|y\|_{L^2([0,t],Y)} + \|u\|_{L^2([0,t],U)} + \|z\|_Z .$$

Further let the operators $\tilde{\mathcal{C}}(t) : H \to X_t$ and $\tilde{\mathcal{D}}(t) : L^2([0,t], U) \to X_t$ be defined by

$$\tilde{\mathcal{C}} = \begin{bmatrix} \mathcal{C}(t) \\ 0 \\ \mathcal{C}_G \end{bmatrix} \quad \text{and} \quad \tilde{\mathcal{D}} = \begin{bmatrix} \mathcal{D}(t) \\ \mathcal{I} \\ \mathcal{D}_G \end{bmatrix},$$

where $\mathcal{I}$ now denotes the identity operator on $L^2([0,t], U)$. Then clearly $\tilde{\mathcal{C}}(t)$ and $\tilde{\mathcal{D}}(t)$ are linear and bounded with adjoints $\tilde{\mathcal{C}}(t)^* : X_t \to H$ and $\tilde{\mathcal{D}}(t)^* : X_t \to L^2([0,t], U)$ given by

$$\tilde{\mathcal{C}}(t)^* = \begin{bmatrix} \mathcal{C}(t)^* & 0 & \mathcal{C}_G^* \end{bmatrix} \quad \text{and} \quad \tilde{\mathcal{D}}(t)^* = \begin{bmatrix} \mathcal{D}(t)^* & \mathcal{I} & \mathcal{D}_G^*. \end{bmatrix}$$

Then it follows that we can factorize the above expression for $P(t)$ as

$$P(t) = \tilde{\mathcal{C}}(t)^* \big(\mathcal{I} + \tilde{\mathcal{D}}(t)\tilde{\mathcal{D}}(t)^*\big)^{-1} \tilde{\mathcal{C}}(t).$$

Hence, the singular value decay of $P(t)$ is the same as that of $\tilde{\mathcal{C}}(t)^*\tilde{\mathcal{C}}(t) \in \mathcal{L}(H)$, i.e. of $\mathcal{C}(t)^*\mathcal{C}(t) + \mathcal{C}_G^*\mathcal{C}_G = \mathcal{C}(t)^*\mathcal{C}(t) + \mathrm{e}^{TA^*}G^*G\mathrm{e}^{TA}$. Applying Weyl's inequality with either the assumption that $\dim Z < \infty$ or that the singular values of e^{tA} decay sufficiently fast yields the second and third assertions, as in the proofs of Theorems 4 and 5. □

Remark 7. *The above theorem extends to the case of a more general cost functional with a coercive weighting term* $\begin{bmatrix} Q & N \\ N^* & R \end{bmatrix}$ *in much the same way as* [27]. *Since* $N = 0$ *in most practical applications and* Q *and* R *may be included in* C *and* B, *respectively, we choose to omit this from the theorem and proof in order to simplify the notation.*

5. Numerical experiments

To demonstrate the validity of the bounds proposed in Theorems 3 to 6 we have performed a few numerical experiments. In all cases, we consider DRE/DLEs arising from LQR problems with the state and output equations given by

$$\dot{x} = Ax + Bu, \quad x(0) = x_0, \tag{13}$$

$$y = Cx, \tag{14}$$

The solution P to the DRE associated with the operators A, B and C yields the optimal input function u^{opt} in feedback form; $u^{\mathrm{opt}}(t) = -B^*P(T-t)x(t)$. It is optimal in the sense that it minimizes the cost functional

$$J(u) = \int_0^T \|y\|_Y^2 + \|u\|_U^2 \, \mathrm{d}t + \|Gy(T)\|_Z^2 .$$

The aim is thus to drive the output y to zero while being mindful of the cost $\|u\|^2$ of doing so. In the extended case mentioned in Remark 7, the weighting factors scale the relative costs of y and u, respectively. When $B = 0$, the solution to the corresponding DLE yields the observability Gramian, an indicator of which states x that can be detected by using only the output y.

In all the following examples we consider the domain $\Omega = [0,1]^2$ to be the unit square, with boundary Γ. We further let the state space be $H = L^2(\Omega)$ except where otherwise noted. We choose $A = \Delta : \mathcal{D}(A) \subset H \to H$ to be the Laplacian. Since we will vary the boundary conditions, its domain will change as well. We can, however, always consider it to be generated by the inner product $a(u,v) = \int_\Omega \nabla u \cdot \nabla v$, where $u, v \in V = \mathcal{D}((-A)^{1/2})$. In the case of homogeneous Dirichlet boundary conditions, we have $\mathcal{D}(A) = H^2 \cap H_0^1(\Omega)$ and $V = H_0^1(\Omega)$. We note that Assumption 1 is satisfied, with the region of analyticity being the entire right halfplane.

Since we cannot investigate the infinite-dimensional case in finite precision arithmetic, a discretization of the equation is required. For the spatial discretization, we have used the finite element method based on the inner product a. For a given mesh size h, we get the finite element space $V_h \subset V \subset H$ and the approximate solution P_h is an operator from V_h to V_h. We may, however, extend it to an operator on H by forming $\mathcal{I}_h P_h \mathcal{P}_h$ where $\mathcal{I}_h : V_h \to H$ denotes the identity operator and $\mathcal{P}_h : H \to V_h$ is the a-orthogonal projection onto the finite element space. For a detailed account of the resulting matrix-valued equations, see e.g. [24, Section 5]. We generate the respective matrices here by using the library FreeFem++ [16], with $P2$ conforming finite elements unless otherwise noted.

Further, since the discretized DLE/DRE are matrix-valued and their solutions are typically dense, it is not feasible to simply transform these into vector-valued ODEs and solve these directly. We use instead the MATLAB package DREsplit[3]

[3]Available from the author via email on request, or from www.tonystillfjord.net.

developed by the author to compute accurate low-rank approximations to the solutions. The reported singular values are thus not exact, but the integration parameters were chosen in such a way that further refining the temporal discretizations has a negligible effect on the end results.

With this said, we want to note that the reported results also provide some insight into how the discretized equations converge to their infinite-dimensional counterparts.

5.1. **Example 1.** We consider first the bounded Lyapunov case by taking the input operator $B = 0$ and letting the output be the mean of the solution. More specifically, we take $Y = \mathbb{R}$ and set $C : H \to Y$, $Cx = \int_\Omega x$. Then clearly $\|Cx\|_\mathbb{R} \leq \|x\|_H$, since Ω is the unit square. We thus have $\beta = 0$ and $\alpha = 0$. Further setting $G = 0$ implies that Assumptions 2 to 4 are satisfied. To complete the specification of A, we choose homogeneous Dirichlet boundary conditions.

We computed the singular values for a number of different spatial discretizations, starting with a grid that has $N = 9$ internal nodes and refining this 6 times. Each refinement roughly halves the mesh size and thus roughly quadruples the number of nodes, leading to meshes with $N = 9, 49, 225, 961, 3969, 16129, 65025$ internal nodes, respectively. Figure 1 shows the computed singular values of the solutions (the $\mathcal{L}(H)$-extended operators, not the matrices) for different spatial discretizations at the final time $T = 0.1$. The curves are ordered in size from bottom to top, i.e. the lowermost curve corresponds to the $N = 9$ discretization, while the topmost corresponds to the $N = 65025$ discretization. We observe that while the initial decay is very much exponential in nature, when we refine the discretization the decay worsens and tends to the exponential square root bound. This is precisely the same behaviour as as seen in the algebraic case in e.g. [15].

5.2. **Example 2.** In the second example, we change the boundary conditions of A to be homogeneous Dirichlet on the left edge Γ_L and homogeneous Neumann on the top and bottom edges Γ_T, Γ_B. On the right edge, Γ_R, we apply a nonhomogeneous Neumann boundary condition, through which we control the system. That is, we set $U = \mathbb{R}$ and define $B : U \to \mathcal{D}(A^*)'$ by $Bu = -(AN\mathbb{1})u$, where the function $\mathbb{1} \in L^2(\Gamma_R)$ is constant equal to 1 everywhere and $N : L^2(\Gamma_R) \to H^{3/2}(\Omega)$ denotes the Neumann operator implicitly defined by $Nv = w$ if $Aw = 0$ in Ω and $\frac{\partial w}{\partial \nu}|_{\Gamma_R} = v$, $\frac{\partial w}{\partial \nu}|_{\Gamma_L \cap \Gamma_T \cap \Gamma_B} = 0$. For further details on this construction, see e.g. [20, Section 3]. That N maps into $H^{3/2}(\Omega)$ follows by [22, Thm. 8.3] and shows that $A^{-\beta}B \in \mathcal{L}(U, H)$ for $\beta = 1/4 + \epsilon$, $\epsilon > 0$.

We note that we could equally well take $U = L^2(\Gamma)$ in the continuous setting and let the input u vary along the whole edge. However, for the numerics we would then have to discretize also this function, leading to one more layer of complexity.

As the output, we again use the mean of the solution over the whole domain Ω, meaning that $\alpha = 0$. We discretize the system in the same way as in Example 1, but because of the three Neumann edges we now have a slightly higher number of degrees of freedom for each level of discretization. The matrices are in this case of size $N = 20, 72, 272, 1056, 4160, 16512, 65792$, respectively.

Figure 2 shows the computed singular values of the solutions at the final time $T = 0.1$. The curves are again ordered in size from coarse (bottom) to fine (top)

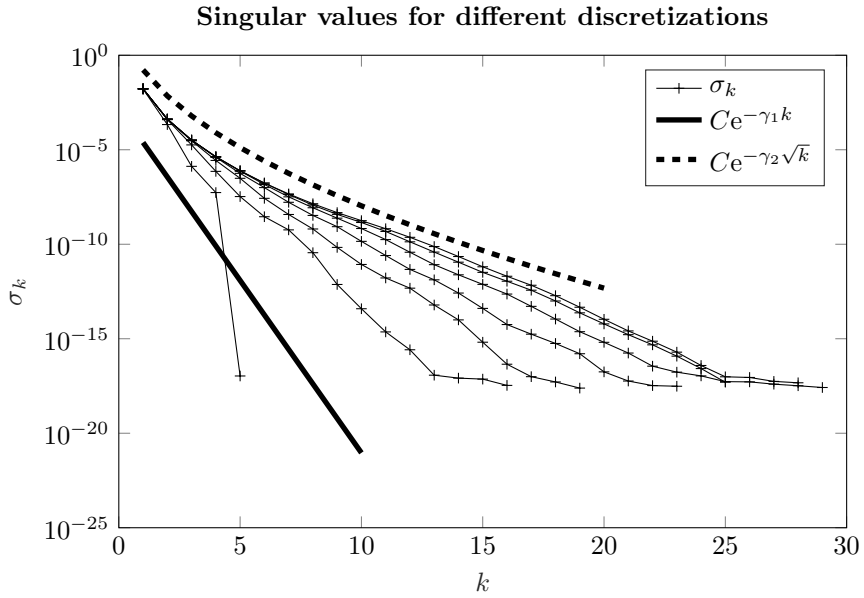

FIGURE 1. The singular values of the solutions computed in Example 1, at the final time $T = 0.1$. They increase monotonically, and thus the lower-most line corresponds to $N = 9$ while the top-most corresponds to $N = 65025$.

discretizations. We note that these results are quite similar to the results in Figure 1, i.e. the input operator does not make the situation worse, as predicted by Theorem 6.

We have additionally plotted the largest singular value of the finest discretized problem as a function of time in Figure 3. We note that it grows roughly as t^1, corresponding well to the factor $t^{1-2\alpha}$ predicted by Theorem 6.

5.3. **Example 3.** Now consider the same setting as in the previous example, but with an unbounded output as well. More precisely, we take $Y = \mathbb{R}$ and define C as the integral of the boundary trace over $\Gamma_T \cap \Gamma_B$:

$$Cx = \int_{\Gamma_T \cap \Gamma_B} x_{|\Gamma}(s) \, \mathrm{d}s.$$

By [22, Theorem 8.3], the map $x \mapsto x_{|\Gamma}$ belongs to $\mathcal{L}(H^{1/2}(\Omega), L^2(\Gamma))$ and hence the map $CA^{-\alpha}$ is bounded for $\alpha = 1/4 + \epsilon$, $\epsilon > 0$.

With the same discretizations as in Example 2, the behaviour of the singular values is similar to when C was bounded. The decay is, however, noticeably slower, as shown in Figure 4. The effect of a larger α can also clearly be seen when plotting the singular values for a specific discretization over time. Figure 5 again shows the largest singular value for the finest discretization. We note that in comparison to Figure 3, the increase is now close to $t^{1/2}$ rather than t^1. Since $\alpha = 1/4$, this is in good agreement with the factor $t^{1-2\alpha}$ predicted by Theorem 6.

5.4. **Example 4.** Let us now consider a situation when the main assumptions are not satisfied. In particular, let us take the same set-up as in Example 3 except for

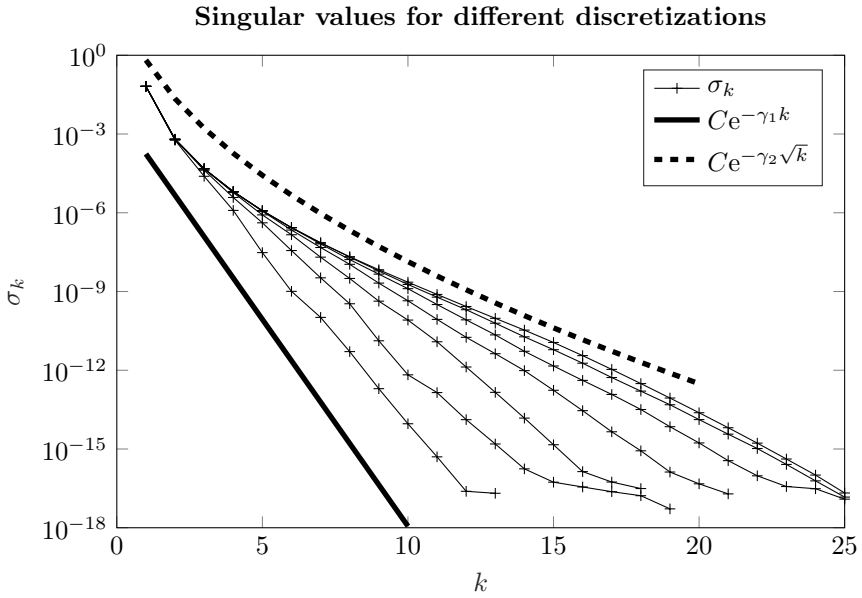

FIGURE 2. The singular values of the solutions computed in Example 2, at the final time $T = 0.1$. They increase monotonically, and thus the lower-most line corresponds to $N = 20$ while the top-most corresponds to $N = 65792$.

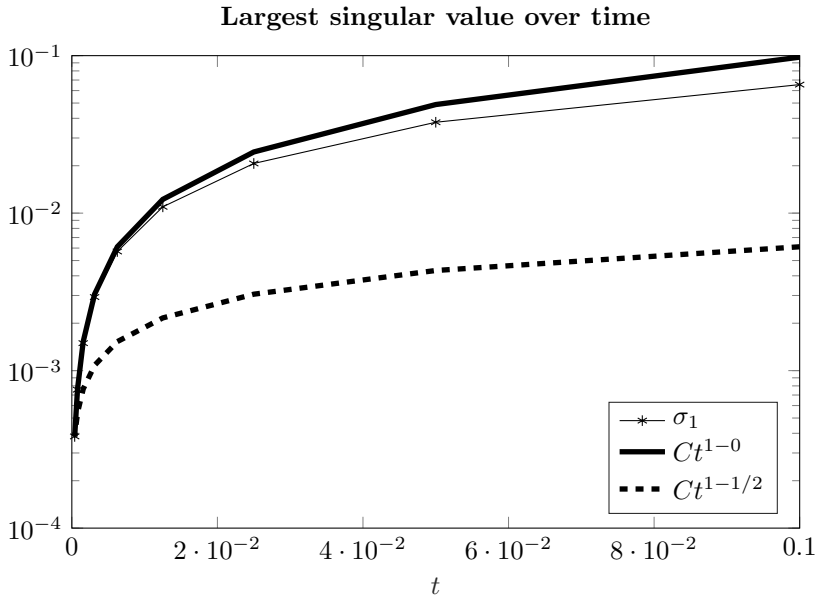

FIGURE 3. The largest singular value of the solution with $N = 65792$ computed in Example 2, plotted over time.

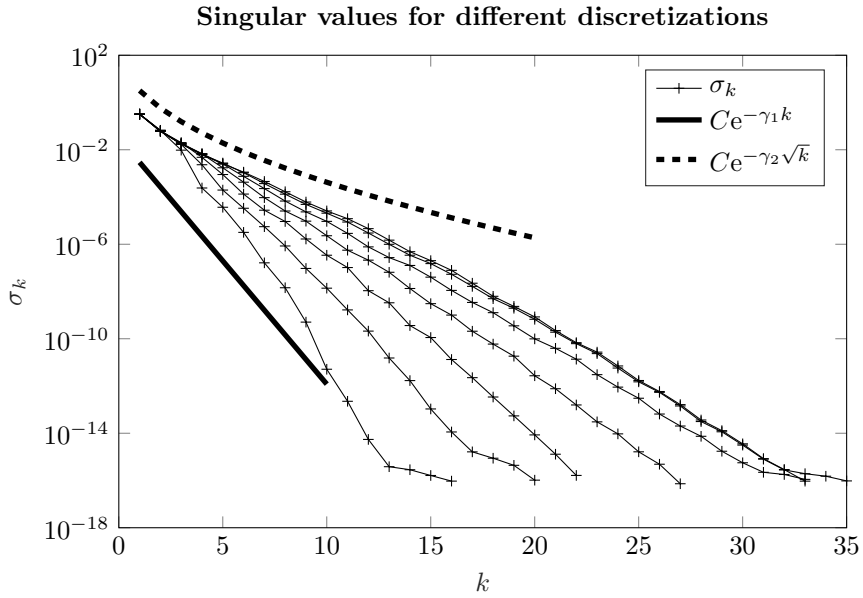

FIGURE 4. The singular values of the solutions computed in Example 3, at the final time $T = 0.1$. They increase monotonically, and thus the lower-most line corresponds to $N = 20$ while the top-most corresponds to $N = 65792$.

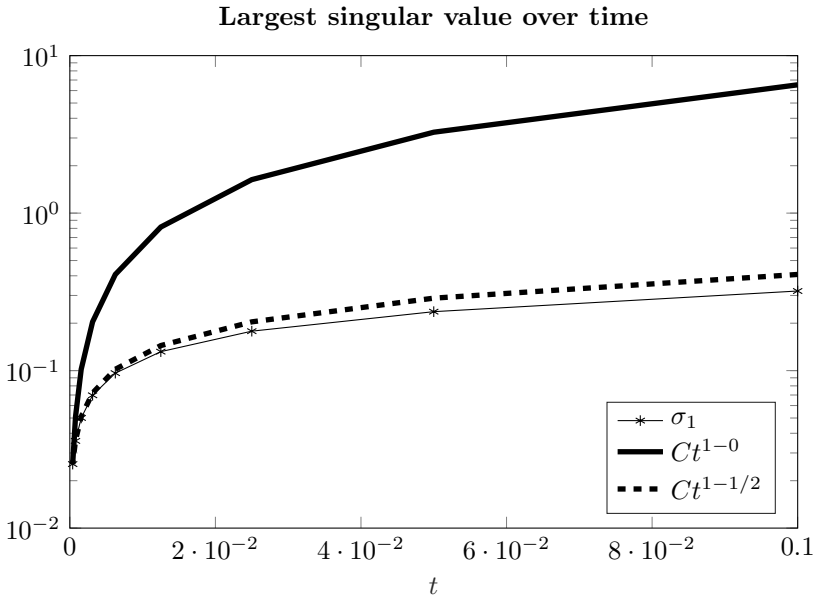

FIGURE 5. The largest singular value of the solution with $N = 65792$ computed in Example 3, plotted over time.

the output operator. We now instead take the trace of the normal derivative:

$$Cx = \int_{\Gamma_T \cap \Gamma_B} \left(\frac{\partial x}{\partial \nu}\right)_{|\Gamma} (s)\, \mathrm{d}s.$$

Again by [22, Theorem 8.3], the map $x \mapsto \left(\frac{\partial x}{\partial \nu}\right)_{|\Gamma}$ belongs to $\mathcal{L}(H^{3/2}(\Omega), L^2(\Gamma))$ and hence the map $CA^{-\alpha}$ is bounded for $\alpha = 3/4 + \epsilon$, $\epsilon > 0$. Since $\alpha > 1/2$, Assumption 3 is not satisfied, and we can in fact not show the existence of a solution $P \in \mathcal{L}(H)$.

This is reflected in the results shown in Figure 6. We have discretized the problem in the same way as previously, and we plot the singular values for the different discretizations like in Figures 1 and 2. In contrast to the previous results, we now see that the singular values keep increasing as we refine the discretization, demonstrating that the singular values of the exact solution are infinite. Thus, while the singular values of a single discretized matrix-valued equation seem to decay exponentially, since the underlying problem is not well posed these "approximations" are nevertheless worthless.

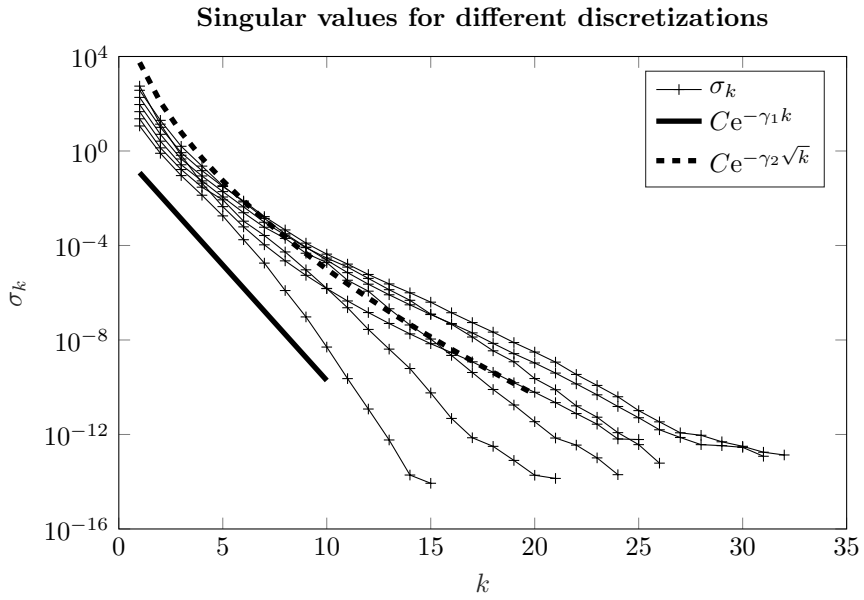

FIGURE 6. The singular values of the solutions computed in Example 4, at the final time $T = 0.1$. They increase (roughly) monotonically, and thus the lower-most line corresponds to $N = 20$ while the top-most corresponds to $N = 65792$. Because the underlying problem is not well-posed, the discretized solutions increase without bound.

5.5. **Example 5.** The situation in the previous Example holds when we use $H = L^2(\Omega)$. By instead selecting a smaller state space H, we decrease the value of α. With $H = \{x \in H^1(\Omega)\ ;\ x_{|\Gamma_L} = 0\}$ and the same operator C we again get

$\alpha = 1/4 + \epsilon$. Since we simultaneously increase β by $1/2$, we set $B = 0$ in this example to comply with Assumption 2.

We note that we now consider the operator A as lifted to H instead of an operator on $L^2(\Omega)$. It still generates an analytic semigroup and Assumption 1 is satisfied. Since the finite-element discretization of the problem is no longer based on $a(u, v)$ but on the corresponding inner product defined on H^1, the resulting problem is similar to a biharmonic equation. This imposes extra regularity requirements on the standard conforming finite element spaces, requiring a high number of nodes [11, p. 286]. In order to avoid this, in this example we employ instead the nonconforming Morley elements [26, 11].

The results are shown in Figure 7. We see that since α is now again less than $1/2$, the singular values behave much like in the previous examples.

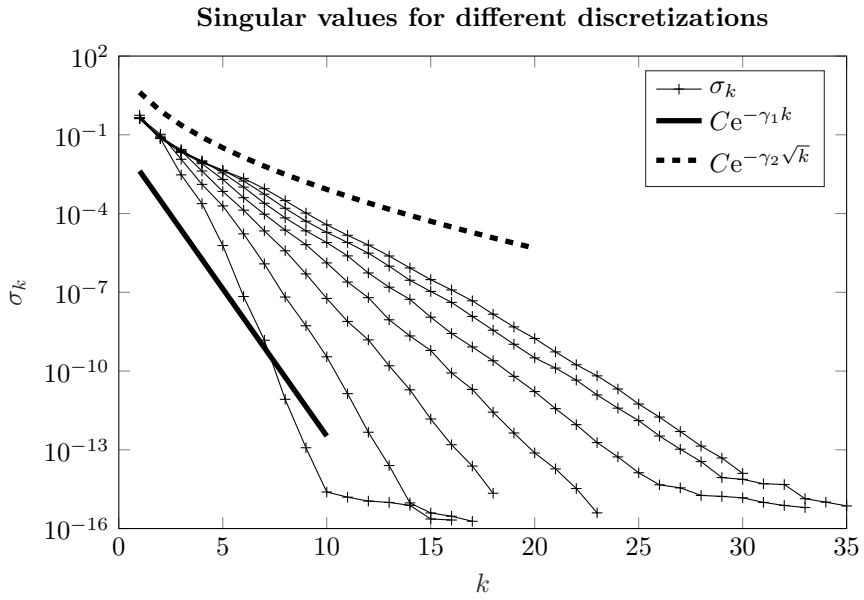

FIGURE 7. The singular values of the solutions computed in Example 5, at the final time $T = 0.1$. They increase monotonically, and thus the lower-most line corresponds to $N = 20$ while the top-most corresponds to $N = 65792$.

6. Conclusions

We have proved bounds for the singular values σ_k of the solutions to DLEs and DREs of the form $\sigma_k \leq Ce^{-\gamma\sqrt{k}}$, extending previous results on algebraic equations to the differential case. This is important, since utilizing the property of low numerical rank is a critical feature in numerical methods for these problems in the large-scale setting. If low numerical rank, i.e. a sufficiently rapid decay of the singular values, can not be guaranteed, these methods never finish, or fail outright. The current work is thus a step on the way to provide practical criteria for when this is to be expected. We say "a step on the way" because while we have given conditions for when exponential square-root decay is to be expected, we have not indicated

how large the constant multiplier in the bound can be. A large value could mean that the numerical rank is too large to be useful in a practical application, even though the decay is $\mathcal{O}(\mathrm{e}^{-\gamma\sqrt{k}})$. However, the size of this constant depends strongly on the properties of the operators A and C, and providing a generally meaningful bound is difficult with current techniques. We therefore leave this question open for future research, but note that the constants arising in our numerical experiments are all of moderate size.

A further interesting unexplored question is how the singular values of the solutions to the spatially discretized matrix-valued problems relate to those of the operator-valued solutions. As noted in the numerical experiments, one often observes exponential decay in the discretized case. When the discretization is refined, the decay rate deteriorates and eventually tends to the exponential square-root bound. The form of this decrease is, however, unclear. While it can be argued that the discretized equations are only steps on the way towards the non-discretized goal (and the author does argue thus), in practical computations we are of course always in the matrix-valued situation. Analysing also this case and providing a connection between the decay rate and the discretization level is therefore both highly interesting and important, but clearly requires a different approach.

7. Acknowledgements

The author is grateful to Mark Opmeer for providing several helpful references.

References

[1] H. Abou-Kandil, G. Freiling, V. Ionescu, and G. Jank, *Matrix Riccati Equations in Control and Systems Theory*, Birkhäuser, Basel, Switzerland, 2003.

[2] A. C. Antoulas, D. C. Sorensen, and Y. Zhou, *On the decay rate of Hankel singular values and related issues*, Syst. Cont. Lett., 46 (2002), pp. 323–342.

[3] T. Başar and P. Bernhard, *H^∞-optimal control and related minimax design problems*, Systems & Control: Foundations & Applications, Birkhäuser Boston, Inc., Boston, MA, second ed., 1995. A dynamic game approach.

[4] J. Baker, M. Embree, and J. Sabino, *Fast singular value decay for Lyapunov solutions with nonnormal coefficients*, arXiv e-prints 1410.8741v1, Cornell University, Oct. 2014. math.NA.

[5] U. Baur, P. Benner, and L. Feng, *Model order reduction for linear and nonlinear systems: A system-theoretic perspective*, Arch. Comput. Methods Eng., 21 (2014), pp. 331–358.

[6] P. Benner and T. Breiten, *Low rank methods for a class of generalized Lyapunov equations and related issues*, Numerische Mathematik, 124 (2013), pp. 441–470.

[7] P. Benner and Z. Bujanović, *On the solution of large-scale algebraic Riccati equations by using low-dimensional invariant subspaces*, Linear Algebra Appl., 488 (2016), pp. 430–459.

[8] P. Benner, P. Kürschner, and J. Saak, *Frequency-limited balanced truncation with low-rank approximations*, SIAM J. Sci. Comput., 38 (2016), pp. A471–A499.

[9] P. Benner, J.-R. Li, and T. Penzl, *Numerical solution of large-scale Lyapunov equations, Riccati equations, and linear-quadratic optimal control problems*, Numer. Lin. Alg. Appl., 15 (2008), pp. 755–777.

[10] A. Bensoussan, G. Da Prato, M. C. Delfour, and S. K. Mitter, *Representation and Control of Infinite Dimensional Systems*, Systems & Control: Foundations & Applications, Birkhäuser, Boston, MA, second ed., 2007.

[11] S. C. Brenner and L. R. Scott, *The mathematical theory of finite element methods*, vol. 15 of Texts in Applied Mathematics, Springer, New York, third ed., 2008.

[12] R. Courant and D. Hilbert, *Methods of mathematical physics. Vol. I*, Interscience Publishers, Inc., New York, N.Y., 1953.

[13] K. Fan, *Maximum properties and inequalities for the eigenvalues of completely continuous operators*, Proc. Nat. Acad. Sci. U.S.A., 37 (1951), pp. 760–766.

[14] W. Gawronski and J.-N. Juang, *Model reduction in limited time and frequency intervals*, Int. J. Syst. Sci., 21 (1990), pp. 349–376.
[15] L. Grubišić and D. Kressner, *On the eigenvalue decay of solutions to operator Lyapunov equations*, Syst. Cont. Lett., 73 (2014), pp. 42–47.
[16] F. Hecht, *New development in freefem++*, J. Numer. Math., 20 (2012), pp. 251–265.
[17] A. Ichikawa and H. Katayama, *Remarks on the time-varying H_∞ Riccati equations*, Syst. Cont. Lett., 37 (1999), pp. 335–345.
[18] P. Kürschner, *Balanced truncation model order reduction in limited time intervals for large systems*, arXiv e-print 1707.02839, Cornell University, 2017. Math.NA.
[19] 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.
[20] I. Lasiecka and R. Triggiani, *Control Theory for Partial Differential Equations: Continuous and Approximation Theories I. Abstract Parabolic Systems*, Cambridge University Press, Cambridge, UK, 2000.
[21] J.-R. Li and J. White, *Low rank solution of Lyapunov equations*, SIAM J. Matrix Anal. Appl., 24 (2002), pp. 260–280.
[22] J.-L. Lions and E. Magenes, *Non-homogeneous boundary value problems and applications. Vol. I*, Springer-Verlag, New York-Heidelberg, 1972. Translated from the French by P. Kenneth, Die Grundlehren der mathematischen Wissenschaften, Band 181.
[23] J. Lund and K. L. Bowers, *Sinc Methods for Quadrature and Differential Equations*, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1992.
[24] A. Målqvist, A. Persson, and T. Stillfjord, *Multiscale differential Riccati equations for linear quadratic regulator problems*, ArXiv e-prints, (2018).
[25] K. M. Mikkola, *Infinite-dimensional linear systems, optimal control and algebraic Riccati equations*, Dissertation, Helsinki University of Technology, Helsinki, Finland, Oct. 2002.
[26] L. S. D. Morley, *The triangular equilibrium element in the solution of plate bending problems*, Aeronaut. Quart., 19 (1968), pp. 149–169.
[27] M. Opmeer, *Decay of singular values of the Gramians of infinite-dimensional systems*, in Proceedings 2015 European Control Conference (ECC), Linz, Austria, 2015, IEEE, pp. 1183–1188.
[28] T. Penzl, *Eigenvalue decay bounds for solutions of Lyapunov equations: the symmetric case*, Syst. Cont. Lett., 40 (2000), pp. 139–144.
[29] I. R. Petersen, V. A. Ugrinovskii, and A. V. Savkin, *Robust Control Design Using H^∞ Methods*, Springer-Verlag, London, UK, 2000.
[30] D. Salamon, *Infinite-dimensional linear systems with unbounded control and observation: a functional analytic approach*, Trans. Amer. Math. Soc., 300 (1987), pp. 383–431.
[31] D. C. Sorensen and Y. Zhou, *Bounds on eigenvalue decay rates and sensitivity of solutions to Lyapunov equations*, Tech. Rep. TR02-07, Dept. of Comp. Appl. Math., Rice University, Houston, TX, June 2002. Available online from http://www.caam.rice.edu/caam/trs/tr02.html#TR02-07.
[32] O. Staffans, *Well-posed linear systems*, vol. 103 of Encyclopedia of Mathematics and its Applications, Cambridge University Press, Cambridge, 2005.
[33] O. J. Staffans, *Quadratic Optimal Control of Regular Well-Posed Linear Systems, with Applications to Parabolic Equations*. Available from http://users.abo.fi/staffans/pdffiles/parabol.pdf, 1997.
[34] F. Stenger, *Integration Formulae Based on the Trapezoidal Formula*, J. Inst. Math. Appl., 12 (1973), pp. 103–114.
[35] ———, *Numerical Methods Based on Sinc and Analytic Functions*, vol. 20 of Springer Series in Computational Mathematics, Springer-Verlag, New York, 1993.
[36] T. Stillfjord, *Low-rank second-order splitting of large-scale differential Riccati equations*, IEEE Trans. Automat. Control, 60 (2015), pp. 2791–2796.
[37] M. Tucsnak and G. Weiss, *Observation and control for operator semigroups*, Birkhäuser Advanced Texts: Basler Lehrbücher. [Birkhäuser Advanced Texts: Basel Textbooks], Birkhäuser Verlag, Basel, 2009.

Max Planck Institute for Dynamics of Complex Technical Systems, Sandtorstr. 1, DE-39106 Magdeburg, Germany
E-mail address: stillfjord@mpi-magdeburg.mpg.de