

Convergence analysis of a full discretization of operator-valued differential Riccati equations

Eskil Hansen Tony Stillfjord Teodor Åberg

September 10, 2026

Abstract

In recent previous work [E. Hansen, T. Stillfjord and T. Åberg, *SIAM J. Numer. Anal.* 64(3) (2026), pp. 957–988], we analyzed the convergence of operator splitting methods applied to operator-valued differential Riccati equations (DRE). In this paper, we extend these results by analyzing the convergence of a full discretization based on finite elements in space and Lie splitting in time. As far as we are aware, this is the first such analysis for DRE. There are very few analyses of temporal discretizations of DRE overall, and none of them have been combined with spatial discretizations. However, it is clearly vital to know when the full discretization converges, since this is what will be used in practical applications. Our main result is that except for logarithmic factors, the method converges with order one in time and order two in space, under fairly weak assumptions on the problem data. This is illustrated by a numerical experiment based on an application in optimal control.

1 Introduction

This paper is a continuation of the work in [17], in which the convergence of splitting schemes applied to operator-valued differential Riccati equations (DRE) was studied. Here, we combine such a temporal discretization with a spatial discretization based on finite elements and analyze the convergence of the resulting full discretization.

A DRE is a semi-linear differential equation of the form

$$\dot{P}(t) = A^*P(t) + P(t)A + Q - P(t)SP(t), \quad t \in (0, T], \quad P(0) = P_0. \quad (1.1)$$

Such equations play a prominent role in optimal control, where their solutions provide the optimal feedback laws in finite-horizon linear quadratic regulator (LQR) problems. We consider the case when the problem data A , Q and S are all operators on a Hilbert space H , which means that also the solution values $P(t)$ are operators on H . This corresponds to optimal control of partial differential equations. To approximate these solutions in practice thus requires both a temporal and a spatial discretization.

Many different methods for DRE have been proposed in the literature, and a brief literature overview is included in the next section. While essentially all spatial discretizations are based on finite elements, a wide variety of temporal discretizations have been suggested. In general, there are very few rigorous analyses of such temporal discretizations in the literature. Spatial discretizations have been analysed more extensively, but mostly for algebraic Riccati equations. We are not aware of any paper which analyses a combination of temporal and spatial discretizations for DRE.

The aim of this paper is therefore to provide a rigorous convergence analysis of such a combination of methods. As in [17], we consider the Lie splitting scheme in time, which iterates between approximating the solutions to the sub-problems $\dot{P} = FP$ and $\dot{P} = GP$, where

$$FP = A^*P + PA + Q \quad \text{and} \quad GP = -PSP.$$

This can be done much more efficiently than directly approximating the solution to (1.1). In space, we consider a large class of discretizations that covers many standard finite element discretizations applicable to problems of the type $Ax = f$.

The analysis uses our temporal convergence results in [17] as a starting point, combined with a new spatial convergence analysis that improves on the state-of-the-art. Let h denote the spatial discretization parameter, let τ be the time step, and let $P_{n,h} \approx P(n\tau)$ denote the fully discretized approximation extended to an operator on H . Then our main result is that for all sufficiently small h and τ ,

$$\|P(n\tau) - P_{n,h}\|_{\mathcal{L}(H)} \leq C(h^2(1 + |\log h|) + \tau(1 + |\log \tau|)).$$

That is, the method is first-order convergent in time and second-order convergent in space, except for logarithmic factors. The result requires some regularity of the operators Q and S , properly defined in Assumption 4. If this assumption is relaxed, a similar result still holds, but with an extra term $h^2(n\tau)^{-1}$. This term reflects the corresponding lack of regularity of the solution P near $t = 0$.

We emphasize that the convergence is measured in the operator norm $\|\cdot\|_{\mathcal{L}(H)}$ for bounded operators on a Hilbert space H . This is much stronger than the pointwise convergence of the type $\|P(n\tau)x - P_{n,h}x\| \rightarrow 0$ for $x \in H$ that is frequently considered, e.g. in [4].

In brief, the analysis approach is as follows:

1. We first compare the exact solution P to the spatially semi-discretized approximation P_h , where the initial condition of the latter is a spatial discretization of P_0 . We bound the spatial error $\|P(t) - P_h(t)\|_{\mathcal{L}(H)}$ by extending the FEM analysis techniques of Crouzeix [11] and Thomée [33] to the present setting of operators of bounded linear operators. The result in [19], which is instead based on optimal control, provide an alternative that requires slightly less regularity but which is also weaker.

2. Again inspired by the approach in the lecture notes by Crouzeix [11], we compare P_h to $\tilde{P}_h$ where $\tilde{P}_h$ is the same spatial semi-discretization of (1.1) as P_h , but with a different, carefully chosen, regularized initial condition $\tilde{P}_{0,h}$.
3. After proving that $\tilde{P}_h$ is uniformly bounded with respect to h , we apply the results in [17] to $\tilde{P}_h(n\tau) - \tilde{P}_{n,h}$, where $\tilde{P}_{n,h}$ is the fully discretized DRE approximation with initial condition $\tilde{P}_{0,h}$.
4. Finally, we bound $\tilde{P}_{n,h} - P_{n,h}$ by following a similar procedure as to what would be used for the corresponding vector-valued elliptic problems.

The rest of the paper is organized in the following manner. Section 3 specifies the details of the setting which we consider, and the main consequences of the assumptions on the problem data A , S , Q and P_0 . The properties of the family of spatial discretizations which fit into our framework are stated in Section 3.1 together with important a priori bounds and convergence results. The existing spatial convergence result from [19] is presented in Section 4, followed by our improved analysis in Section 4.1. Section 5 introduces the full discretization and recalls the important temporal convergence results from [17], then performs the analysis outlined in the previous paragraph. These theoretical results are illustrated by a numerical experiment in Section 6, and concluding remarks are given in Section 7.

2 Literature overview

An operator-valued DRE (1.1) can be discretized in two ways. Either temporally, which yields a sequence of operator-valued algebraic equations, or spatially, which yields a matrix-valued ODE.

There are few temporal discretizations of operator-valued DREs with accompanying error analyses. We are only aware of three. First, [16] provides a rigorous analysis of a splitting scheme. This method, however, does not take full advantage of the structure of the DRE and is therefore less efficient than the method we consider here. The unpublished manuscript [30] considers the same Lie splitting scheme as in this paper, but in a setting of Hilbert-Schmidt operators. This setting is very restrictive and the analysis is therefore only applicable to very few problems. The recent manuscript [17] improves considerably on this by demonstrating that both Lie and Strang splitting converge with the classical orders (except for logarithmic factors) in the usual operator-norm. We mention also [26], which considers only the matrix-valued case, but whose analysis technique inspired the approach in [17].

In terms of spatial discretization of operator-valued DREs, convergence without order is shown in e.g. [28, 4]. As far as we are aware, the first proof of convergence with almost optimal orders in a setting similar to ours is given in [19]. The comprehensive book [22] proves convergence with orders also in more general settings, but only for *algebraic* Riccati equations. The recent works [7, 9]

do consider Galerkin discretizations of DREs. But they do not combine these with temporal discretizations, and additionally work in a setting of compact operators. This is a very strong assumption.

There is a wealth of papers applying different temporal discretizations to matrix-valued DREs, i.e. after a spatial discretization has already been done. We mention BDF methods [2], Rosenbrock methods [3], Peer methods [1], and exponential integrators [23, 8]. The splitting scheme considered in this paper was introduced in [29], see also [31]. The same method but with a different decomposition of the right-hand-side was considered in [25, 26]. Only some of these works feature error analyses, and most of these are focused on estimating the errors introduced by subsequent low-rank-approximation.

We emphasize that none of these works combine spatial and temporal discretizations and provide a rigorous error analysis for the resulting full discretization.

3 Setting

The setting in which we will analyze the method is the same as in [17], but we restate the important notation here for the convenience of the reader. Let H be a Hilbert space over $\mathbb{C}$, and U, Y be two auxiliary Hilbert spaces. In the context of LQR problems, U and Y correspond to the input and output spaces, respectively, while H is the state space. We denote the inner product of a generic Hilbert space X by $(\cdot, \cdot)_X$ and the induced norm by $\|\cdot\|_X$. In both cases the subscript will be omitted if the space is clear from context. Furthermore, $\mathcal{L}(X, Z)$ denotes the Banach space of linear bounded operators from X to another Hilbert space Z , equipped with the usual operator norm. When $X = Z$, this space will be denoted by $\mathcal{L}(X)$. The Hilbert space adjoint of an operator $V : Z \rightarrow X$ is denoted by $V^* : X \rightarrow Z$. The subspace of $\mathcal{L}(X)$ consisting of self-adjoint operators, i.e. those satisfying $(Vx, y) = (x, Vy)$ for all $x, y \in X$, will be denoted by $\Sigma(X)$. Those operators which are additionally non-negative, i.e. where $(Vx, x) \geq 0$ for all $x \in X$, will be denoted by $\Sigma^+(X)$. Finally, we denote the usual supremum norm by

$$\|P\|_\infty = \sup_{t \in (0, T)} \|P(t)\|_H.$$

Throughout the paper, C will denote an arbitrary positive constant which may be different from line to line. It will often depend on the problem data T , $\|P_0\|_{\mathcal{L}(H)}$, $\|Q\|_{\mathcal{L}(H)}$ and $\|S\|_{\mathcal{L}(H)}$, but for brevity we only specify its dependence on other parameters.

We make the following standard assumptions in order to ensure the existence of solutions to (1.1) following the framework presented in [5, Part IV.1].

Assumption 1. *There are operators $B \in \mathcal{L}(U, H)$ and $E \in \mathcal{L}(H, Y)$ such that Q and S can be factorized as $Q = E^*E$ and $S = BB^*$. Further, $P_0 \in \Sigma^+(H)$, and the unbounded operator $A : \mathcal{D}(A) \subset H \rightarrow H$ is the generator of an analytic C_0 -semigroup e^{tA} . Here, $\mathcal{D}(A) = \{x \in H ; Ax \in H\}$.*

Next, for $P \in \Sigma(H)$ define the bilinear map $\tilde{\varphi}_P(\cdot, \cdot) : \mathcal{D}(A) \times \mathcal{D}(A) \rightarrow \mathbb{C}$ by

$$\tilde{\varphi}_P(x, y) = (Px, Ay) + (Ax, Py),$$

and consider the set of P such that it is bounded with respect to the norm of H , that is

$$\mathcal{D}(\mathcal{A}) = \left\{ P \in \Sigma(H) : |\tilde{\varphi}_P(x, y)| \leq C\|x\|\|y\| \text{ for all } x, y \in \mathcal{D}(A) \right\}.$$

For such P , $\tilde{\varphi}_P$ can be extended to an operator $\varphi_P : H \times H \rightarrow \mathbb{C}$, and this gives rise to the operator $\mathcal{A} : \mathcal{D}(\mathcal{A}) \rightarrow \Sigma(H)$ defined by

$$(\mathcal{A}(P)x, y) = \varphi_P(x, y) \quad \text{for all } x, y \in H.$$

This operator is the proper extension of the sum $A^*P + PA$, which is not necessarily in $\Sigma(H)$. In particular, it holds that

$$A^*Px + PAx = \mathcal{A}(P)x \quad \text{for all } x \in \mathcal{D}(A),$$

when $P \in \mathcal{D}(\mathcal{A})$. For further details and a proof of the previous statement, we refer to [5, Part IV.1]. By a slight abuse of notation, we also introduce the operator $\mathcal{A}^*$ following the above approach with A^* instead of A . This is not the adjoint of $\mathcal{A}$ since $\Sigma(H)$ is not a Hilbert space, but it satisfies $APx + PA^*x = \mathcal{A}^*(P)x$ for all $P \in \mathcal{D}(\mathcal{A}^*)$ and $x \in \mathcal{D}(A^*)$.

From now on, we will be working with the equation

$$\dot{P}(t) = \mathcal{A}P(t) + Q - P(t)SP(t), \quad t \in (0, T], \quad P(0) = P_0, \quad (3.1)$$

instead of (1.1). Under Assumption 1, this problem has a unique solution, see [17, Lemmas 2.6 and 2.7] and [5, Proposition IV-1.3.2 and Theorem IV-1.3.1]. To guarantee sufficient regularity of this solution for the temporal discretization to converge with the expected order, we additionally make the following assumption, cf. [17, Assumption 3.1]:

Assumption 2. P_0 is in the domain of $\mathcal{A}$.

Furthermore, (see [5, Section IV-1.3.1]), there is a semigroup $e^{t\mathcal{A}} : \Sigma(H) \rightarrow \Sigma(H)$ associated with $\mathcal{A}$, defined by

$$e^{t\mathcal{A}}P = e^{tA^*}Pe^{tA}.$$

It is important to note that $e^{t\mathcal{A}}$ is not (in general) strongly continuous, which is why we do not say that it is *generated* by $\mathcal{A}$. Nevertheless, we have the pointwise continuity $\lim_{t \rightarrow 0} e^{t\mathcal{A}}Px = Px$ for any $x \in H$ and the important identities

$$\frac{d}{dt}e^{t\mathcal{A}}P = \mathcal{A}e^{t\mathcal{A}}P = e^{t\mathcal{A}}\mathcal{A}P \quad (3.2)$$

hold for $P \in \mathcal{D}(\mathcal{A})$ when $t > 0$. The semigroup is also bounded in the way one would expect, as summarized in the following theorem:

Theorem 3.1. *Let A be the generator of an analytic semigroup. Then there exists a $\omega_0 \geq 0$ such that*

$$\begin{aligned} \|e^{tA}\|_{\mathcal{L}(H)} &\leq Ce^{\omega_0 t}, \quad t \geq 0, \\ \|Ae^{tA}\|_{\mathcal{L}(H)} &\leq \frac{Ce^{\omega_0 t}}{t}, \quad t > 0. \end{aligned}$$

Similarly, with $\mathcal{A}$ defined as above and $M \in \mathcal{L}(H)$,

$$\begin{aligned} \|e^{t\mathcal{A}}M\|_{\mathcal{L}(H)} &\leq Ce^{2\omega_0 t}\|M\|_{\mathcal{L}(H)}, \quad t \geq 0, \\ \|\mathcal{A}e^{t\mathcal{A}}M\|_{\mathcal{L}(H)} &\leq \frac{Ce^{2\omega_0 t}}{t}\|M\|_{\mathcal{L}(H)}, \quad t > 0. \end{aligned}$$

Proof. This follows directly from standard results on analytic semigroups in the same way as Lemma 2.5 in [17]. $\square$

Since we are working on a finite time interval $[0, T]$, we will typically bound $e^{\omega_0 t}$ by $C = e^{\omega_0 T}$. Further, by choosing $\lambda > \omega_0$, the operator $A - \lambda I$ generates an exponentially stable semigroup. That is, it satisfies $\|e^{t(A-\lambda I)}\|_{\mathcal{L}(H)} \leq Ce^{\omega t}$ with $\omega = \omega_0 - \lambda < 0$. This shift λ will be used throughout the manuscript. Note that the corresponding shifted Lyapunov operator is $\mathcal{A} - 2\lambda I$, since $\mathcal{A}$ contains two instances of A . In the above, and throughout the rest of the manuscript, we use the same symbol I to denote the identity operator on different function spaces. It should be clear from the context which one is meant.

3.1 Spatial discretization

We assume that we have a family of finite dimensional subspaces $H_h \subset H$ depending on a discretization parameter $h > 0$, with $\|u\|_{H_h} = \|\pi_h^* u\|_H$ for any u in H_h . Here, the operator $\pi_h : H \rightarrow H_h$ denotes the H -orthogonal projection onto H_h and $\pi_h^* : H_h \rightarrow H$ is the isometric injection operator. The projection π_h gives rise to the operator-projection $\Pi_h : \mathcal{L}(H) \rightarrow \mathcal{L}(H_h)$ given by $\Pi_h M = \pi_h M \pi_h^*$. Its adjoint $\Pi_h^* : \mathcal{L}(H_h) \rightarrow \mathcal{L}(H)$, $\Pi_h^* M = \pi_h^* M \pi_h$ is the isometric injection operator on $\mathcal{L}(H_h)$. Thus, given an operator $P_h \in \mathcal{L}(H_h)$, the natural extension to an operator in $\mathcal{L}(H)$ is given by $\Pi_h^* P_h$, and it holds that

$$\|\Pi_h^* P_h\|_{\mathcal{L}(H)} = \|P_h\|_{\mathcal{L}(H_h)}. \quad (3.3)$$

We note that Π_h and Π_h^* are both bounded operators, and that $\Pi_h \Pi_h^* = I$.

Furthermore, we have a family of operators $A_h : H_h \rightarrow H_h$, which serve as discrete versions of the operator A . Finally, we introduce the “elliptic” operator projection

$$\Theta_h = (\mathcal{A}_h - 2\lambda I)^{-1} \Pi_h (\mathcal{A} - 2\lambda I),$$

which will be used for an auxiliary step in the proof of our improved spatial error bound.

Following [22, Chapter 4], we assume that the approximating spaces and operators fulfill the following bounds:

Assumption 3. Let A fulfill the assumptions given in Assumption 1 and let $h_0 > 0$. The following holds for all $h \leq h_0$. First, the approximating spaces H_h have the property that $\|x - \pi_h x\| \rightarrow 0$ as $h \rightarrow 0$ for all $x \in H$. Furthermore, $A_h \in \mathcal{L}(H_h)$ satisfies

$$\|\Pi_h^* e^{tA_h}\|_{\mathcal{L}(H)} \leq C e^{\omega_0 t}, \quad t \geq 0, \quad (3.4)$$

$$\|\Pi_h^* A_h e^{tA_h}\|_{\mathcal{L}(H)} \leq C \frac{e^{\omega_0 t}}{t}, \quad t > 0. \quad (3.5)$$

Finally, the resolvent of A_h approximates the resolvent of A in the sense that

$$\|(A - \lambda I)^{-1} - \Pi_h^*(A_h - \lambda I)^{-1}\|_{\mathcal{L}(H)} \leq Ch^2. \quad (3.6)$$

Remark 3.1. We note that $(A - \lambda I)^{-1}$ exists as an operator in $\mathcal{L}(H)$ by virtue of $A - \lambda I$ being the generator of an exponentially stable analytic semigroup, see e.g. [27, Theorem 12.31]. That $(A_h - \lambda I)^{-1}$ exists is part of Assumption 3.

Remark 3.2. The inequalities (3.4) and (3.5) are analogous to Assumption (A.1) in Section 4.1.2.1 of [22]. As noted in [22, Remark 4.1.2.2], they are slightly non-standard assumptions that nevertheless typically hold in the present setting where A generates an analytic semigroup. In particular, they are satisfied if A is uniformly elliptic on Ω [21, Lemma 4.2]. The conditions $\|x - \pi_h x\| \rightarrow 0$ and the elliptic error bound (3.6) are standard assumptions. In particular, assume that A arises from a bounded and elliptic bilinear form a via $(Au, v) = a(u, v)$. Further assume that the underlying problem $Ax = f$ is regular (i.e. $\|x\|_{H^2(\Omega)} \leq C\|f\|$), which holds at least when the domain Ω is convex [15, Theorem 9.24] or has a smooth boundary [14, Theorem 9.15, Lemma 9.17]. Then a conforming, piecewise continuous and regular (shape-regular) FEM discretization satisfies (3.6) if Ω is at most 3-dimensional, see [10, Theorem 3.2.5]. We note that since our analysis does not require any inverse inequalities, the mesh is not required to be quasi-uniform. The analysis therefore covers also discretizations in which the mesh is adaptively refined. For the benefit of such a discretization to be visible in the error bound, (3.6) needs to be modified to a local estimate for each finite element K with corresponding diameter $h(K)$ (cf. [10, Theorem 3.1.5] or [6, Theorem 9.1.10]) and then propagated through our entire analysis. For simplicity, readability, and for the sake of not unnecessarily restricting the class of spatial discretizations that are covered, we do not do this here.

Remark 3.3. The restriction $h \leq h_0$ will be enforced throughout the paper, but the choice of h_0 is arbitrary and it has a minor effect on the error constants. It could therefore, e.g., be chosen as the diameter of the computational domain. We will not explicitly write out how it affects the error constants.

A main consequence of these assumptions on A and A_h is that we can also bound the corresponding solution operators in a similar way:

Lemma 3.2 ([22, Proposition 4.1.2.1 (i)]). *Under Assumption 1 and Assumption 3,*

$$\|\Pi_h^* e^{tA_h} - e^{tA}\|_{\mathcal{L}(H)} \leq C e^{\omega_0 t} \frac{h^2}{t}$$

for $t > 0$.

As previously, the constant $e^{\omega_0 t}$ will typically be estimated by $C = e^{\omega_0 T}$ without further notice.

We define $\mathcal{A}_h : \mathcal{L}(H_h) \rightarrow \mathcal{L}(H_h)$ by

$$\mathcal{A}_h U_h = A_h^* U_h + U_h A_h,$$

and similarly $\mathcal{A}_h^* : \mathcal{L}(H_h) \rightarrow \mathcal{L}(H_h)$ is given by

$$\mathcal{A}_h U_h = A_h U_h + U_h A_h^*.$$

The operator $\mathcal{A}_h$ induces the semigroup $e^{t\mathcal{A}_h} : \mathcal{L}(H_h) \rightarrow \mathcal{L}(H_h)$ given by

$$e^{t\mathcal{A}_h} U_h = e^{tA_h^*} U_h e^{tA_h}.$$

For these operators, the bounds of Assumption 3 can be directly extended to yield the following analogue of Theorem 3.1:

Lemma 3.3. *Let Assumption 1 and 3 hold. Then for any $M_h \in \mathcal{L}(H_h)$,*

$$\begin{aligned} \|e^{t\mathcal{A}_h} M_h\|_{\mathcal{L}(H_h)} &\leq C e^{2\omega_0 t} \|M_h\|_{\mathcal{L}(H_h)}, \quad t \geq 0, \\ \|\mathcal{A}_h e^{t\mathcal{A}_h} M_h\|_{\mathcal{L}(H_h)} &\leq \frac{C e^{2\omega_0 t}}{t} \|M_h\|_{\mathcal{L}(H_h)}, \quad t > 0. \end{aligned}$$

The approximation properties have been given in terms of A and A_h , but we are interested in working with the operators $\mathcal{A}$ and $\mathcal{A}_h$ instead. The following two results show that the bound given by (3.6) can be extended to the operators $\mathcal{A}$ and $\mathcal{A}_h$. By the idea in [12, Theorem 4.1.3], we obtain the first result.

Lemma 3.4. *Let Assumptions 1 and 3 hold. The operators $\mathcal{A} - 2\lambda I : \mathcal{D}(\mathcal{A}) \rightarrow \Sigma(H)$ and $\mathcal{A}_h - 2\lambda I : \Sigma(H_h) \rightarrow \Sigma(H_h)$ are both invertible, independent of h . Furthermore, $(\mathcal{A} - 2\lambda I)^{-1} : \Sigma(H) \rightarrow \Sigma(H)$ and $(\mathcal{A}_h - 2\lambda I)^{-1} : \Sigma(H_h) \rightarrow \Sigma(H_h)$ are bounded.*

Proof. Consider first the undiscretized operator, which we abbreviate as $\bar{\mathcal{A}} = \mathcal{A} - 2\lambda I$ for readability. Similarly, we abbreviate $\bar{A} = A - \lambda I$. The shifts do not change the domains of these operators, i.e. $\mathcal{D}(\bar{\mathcal{A}}) = \mathcal{D}(\mathcal{A})$ and $\mathcal{D}(\bar{A}) = \mathcal{D}(A)$. For an arbitrary $M \in \Sigma(H)$ we define P by

$$Pz = \int_0^\infty e^{s\bar{A}^*} M e^{s\bar{A}} z ds$$

for any $z \in H$. This is well-defined, since by Theorem 3.1 the norm of the integrand is bounded by $C^2 e^{2(\omega_0 - \lambda)s} \|M\| \|z\|$. Since $e^{s\bar{A}^*} M e^{s\bar{A}}$ is also symmetric, $P \in \Sigma(H)$.

Now take $x, y \in \mathcal{D}(A)$. We have

$$\begin{aligned} (Px, \bar{A}y) + (\bar{A}x, Py) &= \int_0^\infty (e^{s\bar{A}^*} M e^{s\bar{A}} x, \bar{A}y) ds + \int_0^\infty (\bar{A}x, e^{s\bar{A}^*} M e^{s\bar{A}} y) ds \\ &= \int_0^\infty (e^{s\bar{A}} x, M e^{s\bar{A}} \bar{A}y) ds + \int_0^\infty (e^{s\bar{A}} \bar{A}x, M e^{s\bar{A}} y) ds \\ &= \int_0^\infty \frac{d}{ds} (e^{s\bar{A}} x, M e^{s\bar{A}} y) ds. \end{aligned}$$

Since the integrand satisfies

$$\left\| \frac{d}{ds} (e^{s\bar{A}}x, Me^{s\bar{A}}y) \right\| \leq 2C^2 e^{2(\omega_0 - \lambda)s} \|M\| \|\bar{A}x\| \|\bar{A}y\| \|x\| \|y\|$$

it is integrable on $[0, \infty)$, and we get

$$(Px, \bar{A}y) + (\bar{A}x, Py) = \left[(e^{s\bar{A}}x, Me^{s\bar{A}}y) \right]_0^\infty = -(x, My),$$

since $e^{s\bar{A}}z \rightarrow 0$ as $s \rightarrow \infty$ for any z . Hence

$$\varphi_P(x, y) = -(Mx, y)$$

is continuous, so $P \in \mathcal{D}(\mathcal{A})$ and $\bar{A}P = -M$. Since the choice of M was arbitrary, $\bar{A}$ is surjective. To see that $\bar{A}$ is also injective, take $P_1, P_2 \in \mathcal{D}(\mathcal{A})$ and assume that $\bar{A}(P_1) = \bar{A}(P_2)$, i.e.

$$\varphi_{P_1}(x, y) - \varphi_{P_2}(x, y) = 0 \quad \text{for all } x, y \in H.$$

This implies that

$$(\bar{A}x, (P_1 - P_2)y) + ((P_1 - P_2)x, \bar{A}y) = 0 \quad \text{for all } x, y \in \mathcal{D}(A).$$

If $x_0, y_0 \in \mathcal{D}(A)$ then both the particular choices $x = e^{t\bar{A}}x_0$ and $y = e^{t\bar{A}}y_0$ belong to $\mathcal{D}(A)$ for any $t \geq 0$, so we find that

$$\begin{aligned} 0 &= (\bar{A}e^{t\bar{A}}x_0, (P_1 - P_2)e^{t\bar{A}}y_0) + ((P_1 - P_2)e^{t\bar{A}}x_0, \bar{A}e^{t\bar{A}}y_0) \\ &= \frac{d}{ds} (e^{t\bar{A}}x_0, (P_1 - P_2)e^{t\bar{A}}y_0). \end{aligned}$$

Thus $t \mapsto (e^{t\bar{A}}x_0, (P_1 - P_2)e^{t\bar{A}}y_0)$ is constant. But both $e^{t\bar{A}}x_0$ and $e^{t\bar{A}}y_0$ tend to 0 as $t \rightarrow \infty$, since $e^{t\bar{A}}$ is exponentially stable. Thus the constant must be 0. In particular with $t = 0$,

$$0 = (x_0, (P_1 - P_2)y_0) \quad \text{for all } x_0, y_0 \in \mathcal{D}(A).$$

Since $\mathcal{D}(A)$ is dense in H , this means that $P_1 = P_2$, and thus $\bar{A}$ is invertible. That $\bar{A}^{-1}$ is bounded follows directly from

$$\begin{aligned} \sup_{\|M\|_{\mathcal{L}(H)}=1} \|\bar{A}^{-1}M\|_{\mathcal{L}(H)} &= \sup_{\|M\|_{\mathcal{L}(H)}=1} \left\| - \int_0^\infty e^{s\bar{A}^*} M e^{s\bar{A}} ds \right\|_{\mathcal{L}(H)} \\ &\leq \sup_{\|M\|_{\mathcal{L}(H)}=1} \int_0^\infty C^2 e^{2\omega_0 s} \|M\|_{\mathcal{L}(H)} ds. \end{aligned}$$

By Assumption 3, specifically (3.4), the above arguments are also valid if $\mathcal{A}$ and H are replaced by $\mathcal{A}_h$ and H_h . $\square$

Lemma 3.5. *Let Assumptions 1 and 3 hold. Then there exists a constant C , such that*

$$\|((\mathcal{A} - 2\lambda I)^{-1} - \Pi_h^*(\mathcal{A}_h - 2\lambda I)^{-1}\Pi_h)M\|_{\mathcal{L}(H)} \leq C\|M\|_{\mathcal{L}(H)}h^2(1 + |\log h|).$$

Proof. By the proof of Lemma 3.4, with the same notation, we have that for any $z \in H$

$$\bar{\mathcal{A}}^{-1}Mz = - \int_0^\infty e^{s\bar{\mathcal{A}}^*} M e^{s\bar{\mathcal{A}}} z \, ds,$$

and similarly,

$$\Pi_h^* \bar{\mathcal{A}}_h^{-1} \Pi_h Mz = - \int_0^\infty \Pi_h^* (e^{s\bar{\mathcal{A}}_h^*} \Pi_h M e^{s\bar{\mathcal{A}}_h}) z \, ds.$$

Thus

$$\begin{aligned} & \bar{\mathcal{A}}^{-1}Mz - \Pi_h^* \bar{\mathcal{A}}_h^{-1} \Pi_h Mz \\ &= - \int_0^\infty \left(e^{s\bar{\mathcal{A}}^*} - \Pi_h^* e^{s\bar{\mathcal{A}}_h^*} \right) M e^{s\bar{\mathcal{A}}} z + \Pi_h^* e^{s\bar{\mathcal{A}}_h^*} M \left(e^{s\bar{\mathcal{A}}} z - \Pi_h^* e^{s\bar{\mathcal{A}}_h} z \right) ds \\ &=: \int_0^\infty I(s) z \, ds. \end{aligned}$$

We split the integral into two parts, $I_1(z) = \int_0^{h^2} I(s) z \, ds$ and $I_2(z) = \int_{h^2}^\infty I(s) z \, ds$. By Theorem 3.1 and Assumption 3 both exponentials are uniformly bounded, so $\|I_1(z)\| \leq Ch^2\|M\|_{\mathcal{L}(H)}\|z\|$. By Lemma 3.2, we further get

$$\begin{aligned} \|I_2(z)\| &\leq \int_{h^2}^\infty C e^{\omega_0 s} \frac{h^2}{s} \|M e^{s\bar{\mathcal{A}}} z\| + \|\Pi_h^* e^{s\bar{\mathcal{A}}_h^*} M\|_{\mathcal{L}(H)} C e^{\omega_0 s} \frac{h^2}{s} \|z\| \, ds \\ &\leq Ch^2\|M\|_{\mathcal{L}(H)}\|z\| \int_{h^2}^\infty e^{\omega_0 s} \frac{1}{s} \, ds \\ &\leq Ch^2\|M\|_{\mathcal{L}(H)}\|z\| \left(\int_{h^2}^1 \frac{1}{s} \, ds + \int_1^\infty e^{\omega_0 s} \, ds \right), \end{aligned}$$

so that $\|I_2(z)\| \leq Ch^2(1 + |\log h|)\|M\|_{\mathcal{L}(H)}\|z\|$. Combining the bounds on I_1 and I_2 and taking the supremum over z with $\|z\| = 1$ proves the result. $\square$

For the elliptic projection operator Θ_h we finally have the following result:

Lemma 3.6. *If $M \in \mathcal{D}(\mathcal{A})$, then*

$$\begin{aligned} \|(I - \Pi_h^* \Theta_h)M\|_{\mathcal{L}(H)} &\leq C\|\mathcal{A}M\|_{\mathcal{L}(H)}h^2(1 + |\log h|) \quad \text{and} \\ \|\Pi_h^*(\Theta_h - \Pi_h)M\|_{\mathcal{L}(H)} &\leq C\|\mathcal{A}M\|_{\mathcal{L}(H)}h^2(1 + |\log h|). \end{aligned}$$

Proof. We once again use the notation $\bar{\mathcal{A}} = \mathcal{A} - 2\lambda I$ and $\bar{\mathcal{A}}_h = \mathcal{A}_h - 2\lambda I$. The first bound follows by the identity

$$(I - \Pi_h^* \Theta_h)M = (\bar{\mathcal{A}}^{-1} \bar{\mathcal{A}} - \Pi_h^* \bar{\mathcal{A}}_h^{-1} \Pi_h \bar{\mathcal{A}})M = (\bar{\mathcal{A}}^{-1} - \Pi_h^* \bar{\mathcal{A}}_h^{-1} \Pi_h) \bar{\mathcal{A}}M$$

and Lemma 3.5. The second bound follows by the identity

$$\Pi_h^*(\Theta_h - \Pi_h)M = \Pi_h^*(\Pi_h \Pi_h^* \Theta_h - \Pi_h \bar{\mathcal{A}}^{-1} \bar{\mathcal{A}})M = \Pi_h^* \Pi_h (\Pi_h^* \bar{\mathcal{A}}_h^{-1} \Pi_h - \bar{\mathcal{A}}^{-1}) \bar{\mathcal{A}} M$$

and again employing Lemma 3.5. $\square$

Lemma 3.7. *If $M \in \mathcal{D}(\mathcal{A})$, then*

$$\|(I - \Pi_h^* \Pi_h)M\|_{\mathcal{L}(H)} \leq C \|AM\|_{\mathcal{L}(H)} h^2 (1 + |\log h|).$$

If instead $M \in \mathcal{D}(\mathcal{A}^*)$, then

$$\|(I - \Pi_h^* \Pi_h)M\|_{\mathcal{L}(H)} \leq C \|\mathcal{A}^* M\|_{\mathcal{L}(H)} h^2 (1 + |\log h|).$$

Proof. The first inequality is a direct consequence of combining the two bounds in Lemma 3.6. The second follows in the same way after observing that the statements of Lemma 3.4, 3.5 and 3.6 also hold if $\mathcal{A}$ and $\mathcal{A}_h$ are replaced by $\mathcal{A}^*$ and $\mathcal{A}_h^*$. The proofs are identical except for replacing A and A_h by A^* and A_h^* . $\square$

3.2 Spatially discretized DRE

In order to discretize (3.1) in space, we approximate the operators S and Q by their projections $S_h \in \mathcal{L}(H_h)$ and $Q_h \in \mathcal{L}(H_h)$, defined by

$$S_h = \Pi_h S, \quad Q_h = \Pi_h Q.$$

Lemma 3.8. *Let Assumption 1 hold, then the operators S_h and Q_h are both in $\Sigma^+(H_h)$, and*

$$\begin{aligned} \|S_h\|_{\mathcal{L}(H_h)} &\leq \|S\|_{\mathcal{L}(H)}, \\ \|Q_h\|_{\mathcal{L}(H_h)} &\leq \|Q\|_{\mathcal{L}(H)}. \end{aligned}$$

Proof. The first assertion follows from the fact that

$$(\Pi_h S x, x)_H = (S \pi_h^* x, \pi_h^* x)_H \text{ for all } x \in H_h,$$

and similarly for Q_h . The second follows from $\|\pi_h\|_{\mathcal{L}(H, H_h)} = 1$. $\square$

We will then consider the (spatially) semi-discretized Riccati equation given by

$$\begin{aligned} \dot{P}_h(t) &= \mathcal{A}_h P_h(t) + Q_h - P_h(t) S_h P_h(t), \quad t \in (0, T], \\ P_h(0) &= \Pi_h P_0, \end{aligned} \tag{3.7}$$

for $P_h(t) \in \Sigma^+(H_h)$.

4 Spatial convergence analysis

There exist several convergence results for the problem (3.7). As far as the authors are aware, until now the following was the most sharp:

Theorem 4.1 ([19, Theorem 3.1]). *Let Assumptions 1 and 3 hold. Then the spatial error is bounded by*

$$\|P(t) - \Pi_h^* P_h(t)\|_{\mathcal{L}(H)} \leq Ch^2(t^{-1} + |\log h|).$$

Remark 4.1. The term $\log h$ is removable under further assumptions on P_0 , U and the spaces H_h , see [18]. But since its influence is marginal, we chose to include it in order to improve both the applicability of the results and the readability of the analysis.

The term t^{-1} reflects the typical behaviour of the exact solution P of (3.1) near $t = 0$ when $P_0 \in \Sigma(H)$, i.e. when no extra regularity assumptions are imposed. As was shown in [17], a time discretization based on Lie splitting can not be guaranteed to converge unless $P_0 \in \mathcal{D}(\mathcal{A})$, and such an assumption must therefore also be done for the full discretization studied here. The proof strategy in [19] is based on the connection between Riccati equations and LQR problems rather than the direct approach considered here, and there appears to be no way to use the extra regularity to improve the result. In the next section, we instead use techniques from FEM analysis combined with extra regularity assumptions to remove the singularity. We note that the singularity was recently avoided also in [9], but at the price of working in the setting of compact linear operators. This severely restricts the applicability of the results.

Before proving the improved error bound, we observe that under no extra regularity assumptions we have the following uniform-in- h bound on the spatially discretized DRE solution.

Lemma 4.2. *Let Assumption 1, 2 and 3 hold. Then $\|\Pi_h^* P_h\|_\infty \leq C$ for all $h \leq h_0$.*

Proof. By the same proof as in [5, Lemmas IV.1.2.2–1.2.3, Theorem IV.1.2.1], applied to the finite dimensional problem (3.7), we have that $\|\Pi_h^* P_h\|_\infty$ can be bounded in terms of T , $\|Q_h\|_{\mathcal{L}(H_h)}$, $\|S_h\|_{\mathcal{L}(H_h)}$ and $\|P_{0,h}\|_{\mathcal{L}(H_h)}$. But the terms $\|Q_h\|_{\mathcal{L}(H_h)}$ and $\|S_h\|_{\mathcal{L}(H_h)}$ are uniformly bounded by $\|Q\|_{\mathcal{L}(H)}$ and $\|S\|_{\mathcal{L}(H)}$, respectively, via Lemma 3.8, and $\|P_{0,h}\|_{\mathcal{L}(H_h)} \leq \|P_0\|_{\mathcal{L}(H)}$ by construction. $\square$

4.1 Improved spatial error bound

As a first step, we consider the case when $Q = 0$ and $S = 0$. Modifying [11, Theorem 13], or alternatively the corresponding results in [33], to the current setting yields the convergence result below.

Lemma 4.3. *Let Assumption 1, 2, and 3 hold, then*

$$\|(e^{t\mathcal{A}} - \Pi_h^* e^{t\mathcal{A}_h} \Pi_h) P_0\|_{\mathcal{L}(H)} \leq C \|\mathcal{A} P_0\|_{\mathcal{L}(H)} h^2 (1 + |\log h|).$$

for all $t \in (0, T]$.

Proof. To abbreviate the notation we set $\sigma(h^2) = Ch^2(1 + |\log h|)$. We split the error as

$$(e^{t\mathcal{A}} - \Pi_h^* e^{t\mathcal{A}_h} \Pi_h)P_0 = (I - \Pi_h^* \Theta_h) e^{t\mathcal{A}} P_0 + \Pi_h^* (\Theta_h e^{t\mathcal{A}} - e^{t\mathcal{A}_h} \Pi_h) P_0. \quad (4.1)$$

Lemma 3.6 with $M = e^{t\mathcal{A}} P_0$ then gives the bound

$$\|(I - \Pi_h^* \Theta_h) e^{t\mathcal{A}} P_0\|_{\mathcal{L}(H)} \leq \|\mathcal{A}P_0\|_{\mathcal{L}(H)} \sigma(h^2),$$

and it remains to bound the second term of (4.1). To this end, define

$$U_h(t) = (\Theta_h e^{t\mathcal{A}} - e^{t\mathcal{A}_h} \Pi_h) P_0.$$

For every $\varepsilon > 0$ and $P \in \mathcal{L}(H)$ one has $t \mapsto e^{t\mathcal{A}} P \in C^\infty([\varepsilon, T], \Sigma(H))$, see [17, Section 2]. Since $(\mathcal{A}_h - 2\lambda I)^{-1} \Pi_h$ is bounded, and $\frac{d}{dt} \mathcal{A} e^{t\mathcal{A}} P_0 = \mathcal{A}^2 e^{t\mathcal{A}} P_0$ by (3.2), we obtain

$$\begin{aligned} \frac{d}{dt} \Theta_h e^{t\mathcal{A}} P_0 &= \frac{d}{dt} \left((\mathcal{A}_h - 2\lambda I)^{-1} \Pi_h (\mathcal{A} - 2\lambda I) e^{t\mathcal{A}} P_0 \right) \\ &= (\mathcal{A}_h - 2\lambda I)^{-1} \Pi_h (\mathcal{A} - 2\lambda I) \mathcal{A} e^{t\mathcal{A}} P_0 \\ &= \Theta_h \mathcal{A} e^{t\mathcal{A}} P_0 \end{aligned}$$

for every $t \in [\varepsilon, T]$. Hence,

$$\begin{aligned} \dot{U}_h(t) - \mathcal{A}_h U_h(t) &= (\Theta_h \mathcal{A} e^{t\mathcal{A}} - \mathcal{A}_h e^{t\mathcal{A}_h} \Pi_h) P_0 - (\mathcal{A}_h \Theta_h e^{t\mathcal{A}} - \mathcal{A}_h e^{t\mathcal{A}_h} \Pi_h) P_0 \\ &= (\Theta_h e^{t\mathcal{A}} \mathcal{A} - (\mathcal{A}_h - 2\lambda I) \Theta_h e^{t\mathcal{A}} - 2\lambda \Theta_h e^{t\mathcal{A}}) P_0 \\ &= (\Theta_h e^{t\mathcal{A}} \mathcal{A} - \Pi_h (\mathcal{A} - 2\lambda I) e^{t\mathcal{A}} - 2\lambda \Theta_h e^{t\mathcal{A}}) P_0 \\ &= (\Theta_h - \Pi_h) e^{t\mathcal{A}} (\mathcal{A} - 2\lambda I) P_0 \end{aligned}$$

for all $t \in [\varepsilon, T]$.

Considering $\varepsilon \in (0, 1/2)$ and applying the variation of constants formula to $U_h(t) = U_h(t(1-\varepsilon) + t\varepsilon)$ yields

$$\begin{aligned} U_h(t)x &= e^{t(1-\varepsilon)\mathcal{A}_h} U_h(t\varepsilon)x + \int_{t\varepsilon}^t V_h(s)x \, ds + \int_{t\varepsilon}^{t/2} V_h(s)x \, ds \\ &= (K_1(\varepsilon) + K_2 + K_3(\varepsilon))x, \end{aligned}$$

where $x \in H$ and

$$V_h(s) = e^{(t-s)\mathcal{A}_h} (\Theta_h - \Pi_h) e^{s\mathcal{A}} (\mathcal{A} - 2\lambda I) P_0.$$

To bound the K_1 -term observe that $P_0 \in \mathcal{D}(\mathcal{A})$ implies that $t \mapsto \mathcal{A} e^{t\mathcal{A}} P_0 x$ is continuous in H for $t \in [0, T]$, see [17, Section 2]. The same continuity therefore holds for $t \mapsto U_h(t)x$. This together with Lemma 3.6 gives

$$\lim_{\varepsilon \rightarrow 0^+} \|K_1(\varepsilon)x\| = \|e^{t\mathcal{A}_h} \Pi_h \Pi_h^* (\Theta_h - \Pi_h) P_0 x\| \leq \|\mathcal{A}P_0\|_{\mathcal{L}(H)} \sigma(h^2) \|x\|.$$

For the K_2 -term one has

$$\begin{aligned}\|K_2x\| &\leq \sigma(h^2) \int_{t/2}^t \|\mathcal{A}e^{s\mathcal{A}}(\mathcal{A} - 2\lambda I)P_0\|_{\mathcal{L}(H)}\|x\| \, ds \\ &\leq C\sigma(h^2)\|(\mathcal{A} - 2\lambda I)P_0\|_{\mathcal{L}(H)}\|x\| \int_{t/2}^t \frac{1}{s} \, ds \\ &\leq C(1 + \|\mathcal{A}P_0\|_{\mathcal{L}(H)})\sigma(h^2)\|x\|.\end{aligned}$$

Finally, for K_3 we have for $s \in [t\varepsilon, t/2]$ that

$$V_h(s) = e^{(t-s)\mathcal{A}_h} \frac{d}{ds} \left((\Theta_h - \Pi_h)e^{s\mathcal{A}} \right) - 2\lambda e^{(t-s)\mathcal{A}_h} (\Theta_h - \Pi_h)e^{s\mathcal{A}} P_0.$$

Integration by parts on the first product therefore gives us the reformulation

$$\begin{aligned}K_3(\varepsilon)x &= e^{t/2\mathcal{A}_h}(\Theta_h - \Pi_h)e^{t/2\mathcal{A}}P_0x - e^{t(1-\varepsilon)\mathcal{A}_h}(\Theta_h - \Pi_h)e^{t\varepsilon\mathcal{A}}P_0x \\ &\quad + \int_{t\varepsilon}^{t/2} (\mathcal{A}_h - 2\lambda I)e^{(t-s)\mathcal{A}_h}(\Theta_h - \Pi_h)e^{s\mathcal{A}}P_0x \, ds.\end{aligned}$$

Employing Lemma 3.6 once again gives the K_3 -bound

$$\begin{aligned}\|K_3(\varepsilon)x\| &\leq C\|\mathcal{A}P_0\|_{\mathcal{L}(H)}\sigma(h^2)\|x\| \left(1 + \int_{t\varepsilon}^{t/2} \frac{e^{2\omega_0(t-s)}}{t-s} \, ds\right) \\ &\leq C\|\mathcal{A}P_0\|_{\mathcal{L}(H)}\|x\|\sigma(h^2)(1 + e^{2\omega_0T} \log(2-2\varepsilon)),\end{aligned}$$

for which the limit $\varepsilon \rightarrow 0^+$ is trivial.

The derived bounds on the K_ℓ -terms yield

$$\|\Pi_h^* U_h(t)x\| \leq C \lim_{\varepsilon \rightarrow 0^+} \sum_{\ell=1}^3 \|K_\ell(\varepsilon)x\| \leq C\|\mathcal{A}P_0\|_{\mathcal{L}(H)}\sigma(h^2)\|x\|$$

which implies the sought after convergence result. $\square$

In order to extend the above result to the DRE solutions, we now require the following assumption in addition to Assumption 2:

Assumption 4. *The operator S belongs to $\mathcal{D}(\mathcal{A}^*)$ and the operator Q belongs to $\mathcal{D}(\mathcal{A})$.*

The second condition ensures that the constant term in (3.1) is sufficiently regular, and the first condition ensures that the same holds for the nonlinear term $GP = -PSP$, as demonstrated in the following lemma:

Lemma 4.4. *Let Assumptions 1, 2, 3, and 4 hold. Then $\|\mathcal{A}GP\|_\infty \leq C(1 + \|\mathcal{A}^*S\|_{\mathcal{L}(H)})$.*

Proof. The proof is similar to that of [17, Lemma 3.9]. As H is a Hilbert space, we have the equalities

$$\|\mathcal{A}GP(t)\|_{\mathcal{L}(H)} = \sup_{x,y \in H \setminus \{0\}} \frac{|(\mathcal{A}GP(t)x, y)|}{\|x\|\|y\|} = \sup_{x,y \in H \setminus \{0\}} \frac{|\varphi_{GP(t)}(x, y)|}{\|x\|\|y\|}.$$

Now let $x, y \in \mathcal{D}(A)$. Since $P(t) \in \mathcal{D}(\mathcal{A})$ for $t \in [0, T]$ and $S \in \mathcal{D}(\mathcal{A}^*)$, we have that $P(t)x \in \mathcal{D}(A^*)$ and $SP(t)x \in \mathcal{D}(A^{**}) = \mathcal{D}(A)$. Thus $(\mathcal{A}^*S)P(t)x = (AS + SA^*)P(t)x$, and the two terms are each well-defined. Thus

$$\begin{aligned} -\varphi_{GP(t)}(x, y) &= (P(t)SP(t)x, Ay) + (Ax, P(t)SP(t)y) \\ &\quad + (ASP(t)x, P(t)y) + (P(t)x, ASP(t)y) - ((\mathcal{A}^*S)P(t)x, P(t)y). \end{aligned}$$

With $x' = SP(t)x$ and $y' = SP(t)y$ we therefore have that

$$-\varphi_{GP(t)}(x, y) = \varphi_{P(t)}(x', y) + \varphi_{P(t)}(x, y') - ((\mathcal{A}^*S)P(t)x, P(t)y).$$

Since $P(t) \in \mathcal{D}(\mathcal{A})$ we know that $|\varphi_{P(t)}(x, y)| \leq C\|x\|\|y\|$, which implies that

$$|\varphi_{GP(t)}(x, y)| \leq 2C\|P\|_\infty\|S\|_{\mathcal{L}(H)}\|x\|\|y\| + \|\mathcal{A}^*S\|_{\mathcal{L}(H)}\|P\|_\infty^2\|x\|\|y\|.$$

As $\mathcal{D}(A)$ is dense in H , this inequality extends to $x, y \in H$. The sought after bound then follows as $\|P\|_\infty \leq C$. $\square$

We can now prove the following improved spatial error bound.

Theorem 4.5. *Under the Assumptions 1, 2, 3, and 4 the spatial error is bounded by*

$$\|P(t) - \Pi_h^* P_h(t)\|_{\mathcal{L}(H)} \leq Ch^2(1 + |\log h|)$$

for all $t \in (0, T]$. The error constant C only depends on the problem data $\|\mathcal{A}P_0\|_{\mathcal{L}(H)}$, $\|\mathcal{A}Q\|_{\mathcal{L}(H)}$ and $\|\mathcal{A}^*S\|_{\mathcal{L}(H)}$.

Proof. The variation of constants formula applied to $P(t)$ and $\Pi_h^* P_h(t)$ gives

$$\begin{aligned} P(t) - \Pi_h^* P_h(t) &= (e^{t\mathcal{A}} - \Pi_h^* e^{t\mathcal{A}_h} \Pi_h)P_0 + \int_0^t (e^{(t-s)\mathcal{A}} - \Pi_h^* e^{(t-s)\mathcal{A}_h} \Pi_h)Q \, ds \\ &\quad + \int_0^t (e^{(t-s)\mathcal{A}} - \Pi_h^* e^{(t-s)\mathcal{A}_h} \Pi_h)GP(s) \, ds \\ &\quad + \int_0^t \Pi_h^* e^{(t-s)\mathcal{A}_h} \Pi_h (GP(s) - \Pi_h^* G_h P_h(s)) \, ds. \end{aligned} \tag{4.2}$$

Via the identity

$$\Pi_h^* G_h P_h(s) = (\Pi_h^* P_h(s))(\Pi_h^* \Pi_h S)(\Pi_h^* P_h(s)),$$

we can expand the last integrand in (4.2) as

$$\begin{aligned} GP(s) - \Pi_h^* G_h P_h(s) &= P(s)(S - \Pi_h^* \Pi_h S)P(s) \\ &\quad + (P(s) - \Pi_h^* P_h(s))(\Pi_h^* \Pi_h S)P(s) \\ &\quad + (\Pi_h^* P_h(s))(\Pi_h^* \Pi_h S)(P(s) - \Pi_h^* P_h(s)). \end{aligned}$$

Taking norms in (4.2) and employing Lemmas 3.6, 3.7 and 4.3 gives

$$\begin{aligned} &\|P(t) - \Pi_h^* P_h(t)\|_{\mathcal{L}(H)} \\ &\leq C(\|\mathcal{A}P_0\|_{\mathcal{L}(H)} + \|\mathcal{A}Q\|_{\mathcal{L}(H)} + \|\mathcal{A}GP\|_\infty + \|P\|_\infty^2 \|\mathcal{A}^* S\|_{\mathcal{L}(H)})h^2(1 + |\log h|) \\ &\quad + C(\|P\|_\infty + \|\Pi_h^* P_h\|_\infty) \int_0^t \|P(s) - \Pi_h^* P_h(s)\|_{\mathcal{L}(H)} ds, \end{aligned}$$

and the desired convergence result follows by applying the continuous Grönwall's inequality, see e.g. [13, Proposition 2.1]. Note that the terms $\|\mathcal{A}GP\|_\infty$ and $\|\Pi_h^* P_h\|_\infty$ are bounded, uniformly for all $h \leq h_0$, by Lemmas 4.4 and 4.2, respectively. $\square$

5 Full space-time convergence analysis

In order to arrive at a full discretization of the DRE, we now discretize (3.7) in time using the Lie operator splitting scheme. We formulate it using the operators $F_h, G_h : \mathcal{L}(H_h) \rightarrow \mathcal{L}(H_h)$, which are defined by

$$\begin{aligned} F_h M_h &= \mathcal{A}_h M_h + Q_h \quad \text{and} \\ G_h M_h &= -M_h S_h M_h. \end{aligned}$$

These are analogous to the operators F and G in [17] for the spatially non-discretized case. They give rise to the subproblems

$$\begin{aligned} \dot{M}_h &= F_h M_h, \quad M_h(0) = M_{0,h}, \\ \dot{M}_h &= G_h M_h, \quad M_h(0) = M_{0,h}, \end{aligned}$$

whose solutions we denote by

$$e^{tF_h} M_{0,h} \quad \text{and} \quad e^{tG_h} M_{0,h},$$

respectively, for any $M_{0,h} \in \Sigma^+(H_h)$.

The following auxiliary lemmas correspond to Lemma 2.11 and Lemma 2.12 in [17] and follow directly from applying these results to (3.7) and using the uniform bounds on S_h and Q_h in Lemma 3.8.

Lemma 5.1. *Let Assumption 1 hold and $M_{0,h} \in \Sigma^+(H_h)$. Then $e^{tG_h} M_{0,h} = (I + tM_{0,h}S_h)^{-1} M_{0,h}$, defined on $[0, T]$, fulfills*

$$\|t \mapsto e^{tG_h} M_{0,h}\|_\infty \leq (1 + T\rho\|S\|_{\mathcal{L}(H)})\rho,$$

for all $M_{0,h}$ such that $\|M_{0,h}\|_{\mathcal{L}(H_h)} \leq \rho$. Additionally, the derivatives of this function fulfill

$$\left\| \frac{d^j}{dt^j} \left(t \mapsto e^{tG_h} M_{0,h} \right) \right\|_{\infty} \leq j! \|S\|_{\mathcal{L}(H)}^j (1 + T\rho \|S\|_{\mathcal{L}(H)})^{j+1} \rho^{j+1},$$

for $j = 1, \dots$.

Lemma 5.2. *Let $M_{1,h}, M_{2,h} \in \mathcal{L}(H_h)$ both fulfill $\|M_{i,h}\|_{\mathcal{L}(H_h)} \leq \rho, i = 1, 2$. Then*

$$\|G_h M_{h,1} - G_h M_{h,2}\|_{\mathcal{L}(H_h)} \leq 2\rho \|S\|_{\mathcal{L}(H)} \|M_{h,1} - M_{h,2}\|_{\mathcal{L}(H_h)}.$$

We consider a uniform temporal grid with step size τ , such that $T = N\tau$. The fully discretized scheme is then defined by $\Pi_h^* P_{n,h} \approx \Pi_h^* P_h(n\tau) \approx P(n\tau)$ and

$$P_{n,h} = \mathcal{L}_{h,\tau}^n P_{0,h}, \quad P_{0,h} = P_h(0),$$

with

$$\mathcal{L}_{h,\tau} = e^{\tau F_h} e^{\tau G_h}.$$

Our main result, which will be proved in the rest of this section, is:

Theorem 5.3. *Let Assumption 1, 2 and 3 be fulfilled. Then it holds that*

$$\|P(n\tau) - \Pi_h^* P_{n,h}\|_{\mathcal{L}(H)} \leq C(h^2(n\tau)^{-1} + h^2(1 + |\log h|) + \tau(1 + |\log \tau|))$$

for sufficiently small h and τ . The error constant C only depends on $\|AP_0\|_{\mathcal{L}(H)}$. If in addition Assumption 4 is fulfilled, then it holds that

$$\|P(n\tau) - \Pi_h^* P_{n,h}\|_{\mathcal{L}(H)} \leq C(h^2(1 + |\log h|) + \tau(1 + |\log \tau|))$$

for sufficiently small h and τ . This error constant C only depends on $\|AP_0\|_{\mathcal{L}(H)}$, $\|AQ\|_{\mathcal{L}(H)}$ and $\|A^*S\|_{\mathcal{L}(H)}$.

Consider again the spatially semi-discretized DRE (3.7). To directly apply the temporal convergence results in [17] to this equation, we would have to bound the derivative $\dot{P}_h$ over $(0, T]$, uniformly in h . This is unfortunately not straightforward, since $\|A_h\|_{\mathcal{L}(H_h)} \rightarrow \infty$ as $h \rightarrow 0$ and $\dot{P}_h(t)$ approaches $\mathcal{A}_h \Pi_h P_0$ when $t \rightarrow 0$. For this reason, we introduce the regularized initial condition

$$\tilde{P}_{0,h} := \Theta_h P_0$$

together with the auxiliary equation

$$\begin{aligned} \dot{\tilde{P}}_h(t) &= \mathcal{A}_h \tilde{P}_h(t) + Q_h - \tilde{P}_h(t) S_h \tilde{P}_h(t), \quad t \in (0, T], \\ \tilde{P}_h(0) &= \tilde{P}_{0,h}, \end{aligned} \tag{5.1}$$

where $\tilde{P}_h(t) \in \Sigma^+(H_h)$. Analogously to the full discretization $P_{n,h}$, we also define the auxiliary discretization $\tilde{P}_{n,h}$ by

$$\tilde{P}_{n,h} = \mathcal{L}_{h,\tau}^n \tilde{P}_{0,h}.$$

The initial value $\tilde{P}_{0,h}$ is close to $P_{0,h}$, as shown in the next lemma.

Lemma 5.4. *Let Assumption 1, 2 and 3 hold. Then*

$$\|\Pi_h^* P_{0,h} - \Pi_h^* \tilde{P}_{0,h}\|_{\mathcal{L}(H)} \leq C \|\mathcal{A}P_0\|_{\mathcal{L}(H)} h^2 (1 + |\log h|).$$

Proof. We will use the notation $\bar{\mathcal{A}} = \mathcal{A} - 2\lambda I$ and $\bar{\mathcal{A}}_h = \mathcal{A}_h - 2\lambda I$ that was introduced in Lemma 3.4. First note that both $P_{0,h}$ and $\tilde{P}_{0,h}$ are elements of $\mathcal{L}(H_h)$, and recall that $\Pi_h \Pi_h^*$ is the identity operator on $\mathcal{L}(H_h)$. Further, since $P_0 \in \mathcal{D}(\mathcal{A})$ we may write $P_0 = \bar{\mathcal{A}}^{-1} \bar{\mathcal{A}} P_0$. Thus, recalling that $\Theta_h = \bar{\mathcal{A}}_h^{-1} \Pi_h \bar{\mathcal{A}}$,

$$P_{0,h} - \tilde{P}_{0,h} = \Pi_h \bar{\mathcal{A}}^{-1} \bar{\mathcal{A}} P_0 - \Pi_h \Pi_h^* \bar{\mathcal{A}}_h^{-1} \Pi_h \bar{\mathcal{A}} P_0.$$

By the identity (3.3), Lemma 3.5, with $M = \bar{\mathcal{A}} P_0$, and the fact that Π_h is isometric we therefore get

$$\begin{aligned} \|\Pi_h^* P_{0,h} - \Pi_h^* \tilde{P}_{0,h}\|_{\mathcal{L}(H)} &= \|P_{0,h} - \tilde{P}_{0,h}\|_{\mathcal{L}(H_h)} \\ &\leq \|\bar{\mathcal{A}}^{-1} \bar{\mathcal{A}} P_0 - \Pi_h^* \bar{\mathcal{A}}_h^{-1} \Pi_h \bar{\mathcal{A}} P_0\|_{\mathcal{L}(H)} \\ &\leq C \|\bar{\mathcal{A}} P_0\|_{\mathcal{L}(H)} h^2 (1 + |\log h|). \end{aligned}$$

□

We also have the following bounds, uniform with respect to h . Thus, in particular, $\tilde{\dot{P}}$ is bounded even though $\dot{P}$ is not.

Lemma 5.5. *Let Assumption 1, 2 and 3 hold. Then there exists a parameter $\gamma_0 > 0$ such that*

$$\max\{\|\Pi_h^* P_h\|_\infty, \|\Pi_h^* \tilde{P}_h\|_\infty, \|\Pi_h^* \tilde{\dot{P}}_h\|_\infty\} \leq \gamma_0 \quad (5.2)$$

for all $h \leq h_0$. Here, γ_0 depends on the problem data T , $\|P_0\|_{\mathcal{L}(H)}$, $\|\mathcal{A}P_0\|_{\mathcal{L}(H)}$, $\|Q\|_{\mathcal{L}(H)}$ and $\|S\|_{\mathcal{L}(H)}$.

Proof. The bound on P_h was proven in Lemma 4.2 and the same proof applies also to $\tilde{P}_h$ after substituting (3.7) with (5.1) and observing that Lemma 5.4 gives

$$\|\tilde{P}_{0,h}\|_{\mathcal{L}(H_h)} = \|\Pi_h^* \tilde{P}_{0,h}\|_{\mathcal{L}(H)} \leq \|P_0\|_{\mathcal{L}(H)} + Ch_0^2 (1 + |\log h_0|) \|\mathcal{A}P_0\|_{\mathcal{L}(H)}.$$

Now consider the derivative of $\tilde{P}_h$. For brevity, let $V_h = \tilde{\dot{P}}_h$. We have that

$$\begin{aligned} \dot{V}_h(t) &= \mathcal{A}_h V_h(t) - V_h(t) S_h \tilde{P}_h(t) - \tilde{P}_h(t) S_h V_h(t), \quad t \in (0, T], \\ V_h(0) &= \mathcal{A}_h \tilde{P}_{0,h} + Q_h - \tilde{P}_{0,h} S_h \tilde{P}_{0,h}, \end{aligned}$$

where

$$\mathcal{A}_h \tilde{P}_{0,h} = \mathcal{A}_h \Theta_h P_0 = \Pi_h (\mathcal{A} - 2\lambda I) P_0 + 2\lambda \Theta_h P_0$$

and the above bound on $\|\tilde{P}_{0,h}\|_{\mathcal{L}(H_h)}$ means that

$$\|V_h(0)\|_{\mathcal{L}(H_h)} \leq C(1 + \|\mathcal{A}P_0\|_{\mathcal{L}(H)}).$$

By the variation of constants formula we can represent this as

$$V_h(t) = e^{t\mathcal{A}_h} V_h(0) - \int_0^t e^{(t-s)\mathcal{A}_h} (\tilde{P}_h(s) S_h V(s) + V(s) S_h \tilde{P}(s)) ds.$$

Taking the norm results in

$$\begin{aligned} \|V_h(t)\|_{\mathcal{L}(H_h)} &\leq \|e^{\mathcal{A}_h t}\|_{\mathcal{L}(H_h)} \|V_h(0)\|_{\mathcal{L}(H_h)} \\ &\quad + 2 \int_0^t \|e^{\mathcal{A}_h(t-s)}\|_{\mathcal{L}(H_h)} \|S\|_{\mathcal{L}(H)} \|\Pi_h^* \tilde{P}_h\|_{\infty} \|V_h(s)\|_{\mathcal{L}(H_h)} ds \\ &\leq C(1 + \|\mathcal{A}P_0\|_{\mathcal{L}(H)}) + C\|S\|_{\mathcal{L}(H)} \|\Pi_h^* \tilde{P}_h\|_{\infty} \int_0^t \|V_h(s)\|_{\mathcal{L}(H_h)} ds. \end{aligned}$$

The continuous Grönwall's inequality, see e.g. [13, Proposition 2.1], then yields that

$$\|\Pi_h^* \dot{\tilde{P}}(t)\|_{\mathcal{L}(H)} = \|V_h(t)\|_{\mathcal{L}(H_h)} \leq C(1 + \|\mathcal{A}P_0\|_{\mathcal{L}(H)}) e^{C\|S\|_{\mathcal{L}(H)} T \|\Pi_h^* \tilde{P}_h\|_{\infty}}.$$

As the right-hand-side does not depend on t and $\|\Pi_h^* \tilde{P}_h\|_{\infty}$ is uniformly bounded with respect to h , we have also derived a uniform bound for $\|\Pi_h^* \dot{\tilde{P}}\|_{\infty}$. $\square$

Remark 5.1. The value γ_0 defined in Lemma 5.5 will be used throughout the rest of the analysis without further reference to the lemma.

We can now rewrite the full error in the following way:

$$\begin{aligned} \|P(n\tau) - \Pi_h^* P_{n,h}\|_{\mathcal{L}(H)} &\leq \|P(n\tau) - \Pi_h^* P_h(n\tau)\|_{\mathcal{L}(H)} \\ &\quad + \|\Pi_h^* (P_h(n\tau) - \tilde{P}_h(n\tau))\|_{\mathcal{L}(H)} \\ &\quad + \|\Pi_h^* (\tilde{P}_h(n\tau) - \tilde{P}_{n,h})\|_{\mathcal{L}(H)} \\ &\quad + \|\Pi_h^* (\tilde{P}_{n,h} - P_{n,h})\|_{\mathcal{L}(H)} \\ &=: E_1^n + E_2^n + E_3^n + E_4^n. \end{aligned} \tag{5.3}$$

This formalises the outline of the analysis given in Section 1; the i -th item in the list corresponds to E_i^n . The first term E_1^n is directly bounded in terms of h by Theorem 4.1 or Theorem 4.5. The following lemmas bound the remaining terms.

Lemma 5.6. *Let Assumption 1, 2 and 3 be fulfilled. Then*

$$E_2^n \leq C(\gamma_0) h^2 (1 + |\log h|)$$

for all $n = 0, \dots, N$, and $h \leq h_0$.

Proof. The operator $W_h = P_h - \tilde{P}_h$ fulfills the equation

$$\begin{aligned} \dot{W}_h &= \mathcal{A}_h W_h - P_h S_h P_h + \tilde{P}_h S_h \tilde{P}_h = \mathcal{A}_h W_h - W_h S_h P_h - \tilde{P}_h S_h W_h, t \in (0, T], \\ W_h(0) &= P_{0,h} - \tilde{P}_{0,h}, \end{aligned}$$

and we have the representation

$$W_h(t) = e^{t\mathcal{A}_h}(P_{0,h} - \tilde{P}_{0,h}) - \int_0^t e^{(t-s)\mathcal{A}_h}(W_h(s)S_hP_h(s) + \tilde{P}_h(s)S_hW_h(s))ds.$$

By Lemmas 3.3, 5.4 and 5.5 we have the bound

$$\|W_h(t)\|_{\mathcal{L}(H_h)} \leq C\|\mathcal{A}P_0\|_{\mathcal{L}(H)}h^2(1 + |\log h|) + C\gamma_0 \int_0^t \|W_h(s)\|_{\mathcal{L}(H_h)}ds.$$

The continuous Grönwall's inequality once again implies that

$$\|\Pi_h^*(P_h(t) - \tilde{P}_h(t))\|_{\mathcal{L}(H)} = \|W_h(t)\|_{\mathcal{L}(H_h)} \leq C\|\mathcal{A}P_0\|_{\mathcal{L}(H)}h^2(1 + |\log h|)e^{C\gamma_0 T}$$

for all $t \in [0, T]$, and the desired bound follows. $\square$

Lemma 5.7. *Let Assumption 1, 2, and 3 hold. Then there exist a constant $\tau_0 > 0$, such that*

$$\|\Pi_h^*\tilde{P}_{n,h}\|_{\mathcal{L}(H)} \leq \gamma_0 + 1 \quad \text{and} \quad E_3^n \leq C(\gamma_0)\tau(1 + |\log \tau|)$$

for all $n = 0, \dots, N$, $h \leq h_0$, and $\tau \leq \tau_0$. Here, τ_0 depends on the parameter γ_0 .

Proof. For a fixed h , the result follows from [17, Theorem 3.6] applied to (5.1). What needs to be verified is that there are h -independent constants τ_0 and $C(\gamma_0)$. The crucial element of the proof of [17, Theorem 3.6] is to choose a parameter γ such that

$$\max\{\|\Pi_h^*\tilde{P}_h\|_\infty, \|\Pi_h^*\dot{\tilde{P}}_h\|_\infty\} \leq \gamma. \quad (5.4)$$

Lemma 5.5 enables the choice $\gamma = \gamma_0$, which implies that the bound (5.4) holds for all $h \leq h_0$. The constants also depend on $\|S_h\|_{\mathcal{L}(H_h)}$, $\|Q_h\|_{\mathcal{L}(H_h)}$, $\|e^{\tau\mathcal{A}_h}\|_{\mathcal{L}(H_h)}$, and $\|\mathcal{A}_he^{\tau\mathcal{A}_h}\|_{\mathcal{L}(H_h)}$. The first two of these terms are uniformly bounded by Lemma 3.8, and for fixed τ the last two are uniformly bounded in h by Lemma 3.3. Thus no h -dependent bounds are required in the derivation of the constants. $\square$

It remains to bound E_4^n . This requires an additional induction argument, as $\mathcal{L}_{\tau,h}^n$ is not globally Lipschitz continuous.

Lemma 5.8. *Let Assumption 1, 2 and 3 hold. If $\|\Pi_h^*P_{k,h}\|_{\mathcal{L}(H)} \leq \gamma_0 + 2$ for $k = 0, \dots, n$, then*

$$E_4^{n+1} \leq D_0(h^2(1 + |\log h|) + \tau)$$

for all $h \leq h_0$ and $\tau \leq \tau_0$. Here, D_0 only depends on the parameter γ_0 .

Proof. For a generic initial condition $M_{0,h} \in \Sigma^+(H_h)$, the application of the operator $\mathcal{L}_{h,\tau}$ can be formulated in the following way:

$$\begin{aligned} \mathcal{L}_{h,\tau}M_{0,h} &= e^{\tau\mathcal{A}_h}M_{0,h} + \int_0^\tau e^{(\tau-s)\mathcal{A}_h}Q_h ds + \tau e^{\tau\mathcal{A}_h}G_hM_{0,h} \\ &\quad + \int_0^\tau (\tau-s)e^{\tau\mathcal{A}_h}\frac{d^2}{ds^2}e^{sG_h}M_{0,h} ds. \end{aligned} \quad (5.5)$$

By applying (5.5) to $P_{0,h}$ and $\tilde{P}_{0,h}$ we can thus recursively express the $(n+1)$ th error in terms of the errors at previous time steps as

$$\begin{aligned} P_{n+1,h} - \tilde{P}_{n+1,h} &= e^{n\tau\mathcal{A}_h} (P_{0,h} - \tilde{P}_{0,h}) + \sum_{k=0}^n e^{(n-k)\tau\mathcal{A}_h} \tau (G_h P_{k,h} - G_h \tilde{P}_{k,h}) \\ &\quad + \sum_{k=0}^n e^{(n-k)\tau\mathcal{A}_h} \int_0^\tau (\tau-s) \left(\frac{d^2}{ds^2} e^{sG_h} P_{k,h} - \frac{d^2}{ds^2} e^{sG_h} \tilde{P}_{k,h} \right) ds. \end{aligned}$$

Taking the norm and using Lemmas 3.3, 5.1, 5.2, 5.4, and 5.7 together with the hypothesis $\|\Pi_h^* P_{k,h}\|_{\mathcal{L}(H)} \leq \gamma_0 + 2$ results in

$$E_4^{n+1} \leq C \|\mathcal{A}P_0\|_{\mathcal{L}(H)} h^2 (1 + |\log h|) + C(\gamma_0) \tau \sum_{k=0}^n E_4^k + C(\gamma_0) \tau.$$

Thus, the desired bound follows by applying a discrete Grönwall inequality, see e.g. [13, Proposition 4.1]. $\square$

Lemma 5.9. *Let Assumption 1, 2 and 3. Then there exist constants $h^* \leq h_0$ and $\tau^* \leq \tau_0$ such that*

$$E_4^n \leq D(h^2(1 + |\log h|) + \tau),$$

for all $n = 0, \dots, N$, $h \leq h^*$, and $\tau \leq \tau^*$. Here, D only depends on the parameter γ_0 .

Proof. By Lemma 5.4, we can choose $h^* \leq h_0$ and $\tau^* \leq \tau_0$ such that

$$\|\Pi_h^* \tilde{P}_{0,h} - P_0\|_{\mathcal{L}(H)} \leq D(h^2(1 + |\log h|) + \tau) \leq 1$$

for all $h \leq h^*$ and $\tau \leq \tau^*$, where $D \geq D_0$. Let $h \leq h^*$ and $\tau \leq \tau^*$ and assume that $E_4^k \leq D(h^2(1 + |\log h|) + \tau)$ for $k = 0, \dots, n$. Lemma 5.7 then gives

$$\|\Pi_h^* P_{k,h}\|_{\mathcal{L}(H)} \leq \|\Pi_h^* \tilde{P}_{k,h}\|_{\mathcal{L}(H)} + E_4^k \leq (\gamma_0 + 1) + 1,$$

and Lemma 5.8 implies that $E_4^{n+1} \leq D(h^2(1 + |\log h|) + \tau)$. As $E_4^0 = \|\Pi_h^* \tilde{P}_{0,h} - P_0\|_{\mathcal{L}(H)}$, the sought after bound follows by induction. $\square$

Combining Theorem 4.1 or Theorem 4.5 with Lemmas 5.6, 5.7, and 5.9 to bound each of the four terms in (5.3) now proves the main result, Theorem 5.3.

6 Numerical experiments

We illustrate the theoretical results by applying the full discretization to a DRE arising from a LQR problem. We use the same setup as in [17], but generalized from 1D to 2D. Thus $H = L^2(\Omega)$ with $\Omega = (0, 1) \times (0, 1)$. We choose A to be the

Laplacian with periodic boundary conditions, i.e. $A = \Delta$ and $\mathcal{D}(A) = H_{\text{per}}^2(\Omega)$. This operator generates an analytic semigroup, see Appendix A. With $Ex = \int_{(1/2,1) \times (0,1)} x$, the problem is a controlled heat equation on the unit square, where we observe the mean of the state x on one half of the domain Ω . We let the control consist of a single input signal which scales a function on Ω . Thus $B : \mathbb{R} \rightarrow H$ with $Bu = u\xi$ where we let $\xi \in H^2(\Omega)$. Then $B^*v = (v, \xi)_H$ so that $S = BB^* : H \rightarrow H$ satisfies $Sv = BB^*v = (\xi, v)_H \xi$. We construct the initial condition $P_0 = L^*L : H \rightarrow H$ in a similar way, by setting $L^*x = x\zeta$, where $\zeta \in H^2(\Omega)$. By construction, Assumption 1 is now satisfied. Since L maps into $H^2(\Omega)$, Assumption 2 is also satisfied.

For the spatial discretization, we use the finite element method with piecewise linear basis functions φ_j , $j = 1, \dots, N_h$. Thus $H_h := \text{span}(\{\varphi_j\}_{j=1}^{N_h})$, and the operators A_h are defined via $(A_h u_h, v_h)_{H_h} = -(\nabla u_h, \nabla v_h)_H$ for all $u_h, v_h \in H_h$. We use a simple grid where Ω is first cut into equi-sized squares which are then subdivided into triangles. With N_x squares in each cardinal direction, we have $\mathcal{O}(N_x^2)$ degrees of freedom in total.

Since the domain is convex, we have the bound $\|\pi_h x - x\| \leq Ch^2 \|Ax\|$, see e.g. [20, Inequality (5.39)]. Combined with the dense embedding of $\mathcal{D}(A)$ in H , this shows that $\|\pi_h x - x\| \rightarrow 0$ as $h \rightarrow 0$ for any $x \in H$. Inequality (3.6) follows from e.g. [20, Theorem 5.4 and (5.25)]. Further, the inequalities (3.4) and (3.5) follows as in the proof of Lemma 4.2 in [21], case (H3)(iii). (Note that a general assumption in this paper is that an inverse inequality holds, but this property is not required for this result.) In conclusion, also Assumption 3 is fulfilled.

As outlined in [17], the approximation $P_h(t)$ can be represented by a matrix $\mathbf{P}(t) \in \mathbb{R}^{N_h \times N_h}$ which solves a matrix-valued DRE

$$\dot{\mathbf{M}}\mathbf{P}\mathbf{M} = \mathbf{M}\mathbf{P}\mathbf{A} + \mathbf{A}^T\mathbf{P}\mathbf{M} + \mathbf{E}^T\mathbf{E} - \mathbf{M}\mathbf{P}\mathbf{B}\mathbf{B}^T\mathbf{P}\mathbf{M},$$

where $\mathbf{M}$ is the mass matrix, $\mathbf{A}$ is the stiffness matrix and $\mathbf{E}$ and $\mathbf{B}$ are the matrix representations of E and B , respectively. We get the matrix representation of $P_{n,h}$ by applying the Lie splitting scheme to this equation. Here, we choose $N_x = 2^k$, $k = 2, \dots, 8$, corresponding to approximations $\mathbf{P}(t) \in \mathbb{R}^{N_x^2 \times N_x^2}$ and to $h = 2^{-k}$. For each N_x , we try different numbers of time steps, $N_t = 2^j$, $j = 1, \dots, 13$, corresponding to $\tau = 2^{-j}$.

Since an exact solution is not available and cannot be “manufactured” without destroying the structure of the DRE, we use the approximation with $N_x = 2^8$ and $N_t = 2^{13}$ as a reference solution $\{P_n^{\text{ref}}\}_{n=1}^{2^{13}}$. When comparing it to $\{P_{n,h}\}_{n=1}^{2^j}$, we restrict it to $\{P_{2^j-13n}^{\text{ref}}\}_{n=1}^{2^j}$ so that the approximations are given on the same temporal grid with 2^j points. In order for the spatial grids also agree, we extend $P_{n,h}$ to $\text{Id}_h^{\mathfrak{h}} P_{n,h} (\text{Id}_h^{\mathfrak{h}})^*$, where $\mathfrak{h} = 2^{-8}$ and $\text{Id}_h^{\mathfrak{h}}$ is the injection operator from H_h to $H_{\mathfrak{h}}$. In total, we then compute the relative discretized $\|\cdot\|_{\infty}$ -errors

$$\text{err}_{\tau,h} = \frac{\max_{n=1,\dots,2^j} \|\text{Id}_{\mathfrak{h}} (\text{Id}_h^{\mathfrak{h}} P_{n,h} (\text{Id}_h^{\mathfrak{h}})^* - P_{2^j-13n}^{\text{ref}}) (\text{Id}_{\mathfrak{h}})^*\|_{\mathcal{L}(H)}}{\max_{n=1,\dots,2^j} \|\text{Id}_{\mathfrak{h}} P_{2^j-13n}^{\text{ref}} (\text{Id}_{\mathfrak{h}})^*\|_{\mathcal{L}(H)}}$$

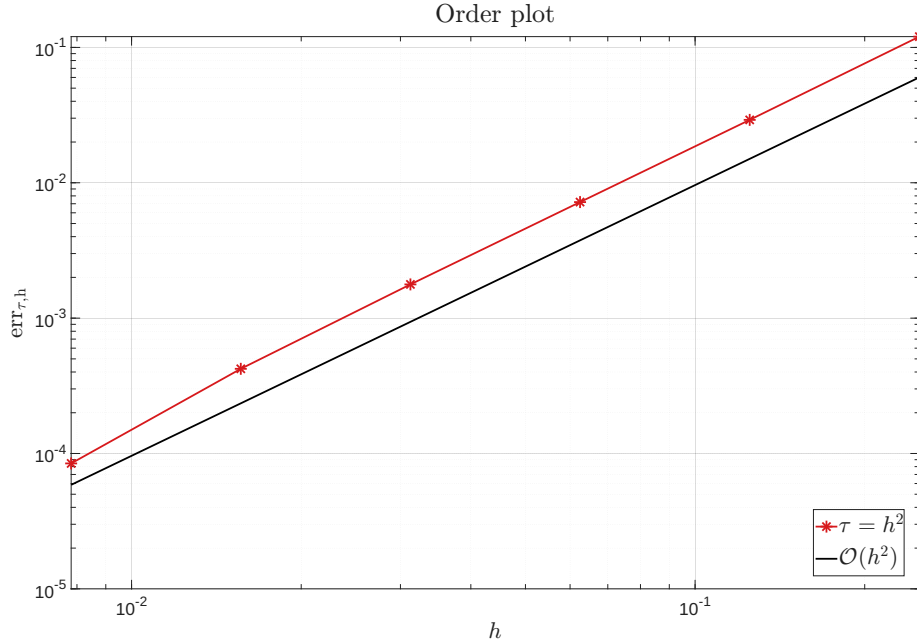


Figure 1: The errors $\text{err}_{\tau,h}$ plotted versus h when $\tau = h^2$. The errors converge like $\mathcal{O}(h^2)$ as expected.

when $\tau = 2^{-j}$. The numerical computation of the $\|\cdot\|_{\mathcal{L}(H)}$ -norms is done as described in [24, Section 5.1]. It requires a one-time computation of a Cholesky-factorization of the mass matrix. These factors, as well as all other critical parameters and matrices, are available at [32]. To compute the approximations $P_{n,h}$ we use the MATLAB library DREsplit¹, which provides an implementation of Lie splitting in low-rank form. The low-rank tolerance was set to 10^{-16} so that the low-rank approximation did not influence the results.

From Theorem 5.3, we expect that $\text{err}_{\tau,h}$ behaves as $\mathcal{O}(h^2 + \tau)$ except for logarithmic factors. This is indeed what we observe in Figure 1. In Figure 2, we have additionally plotted the errors against τ , with one curve for each h . Here, we can see how the error decreases like $\mathcal{O}(\tau)$ when the temporal errors are dominant, but eventually stagnates at the level of the spatial error. Conversely, Figure 3 shows how the error decreases like $\mathcal{O}(h^2)$ until stagnating due to the temporal error.

7 Conclusions

We have proved convergence of a full discretization of operator-valued DRE with optimal orders, except for logarithmic factors. That is, they coincide with the

¹Available from the corresponding author on request, or from www.tonystillfjord.net.

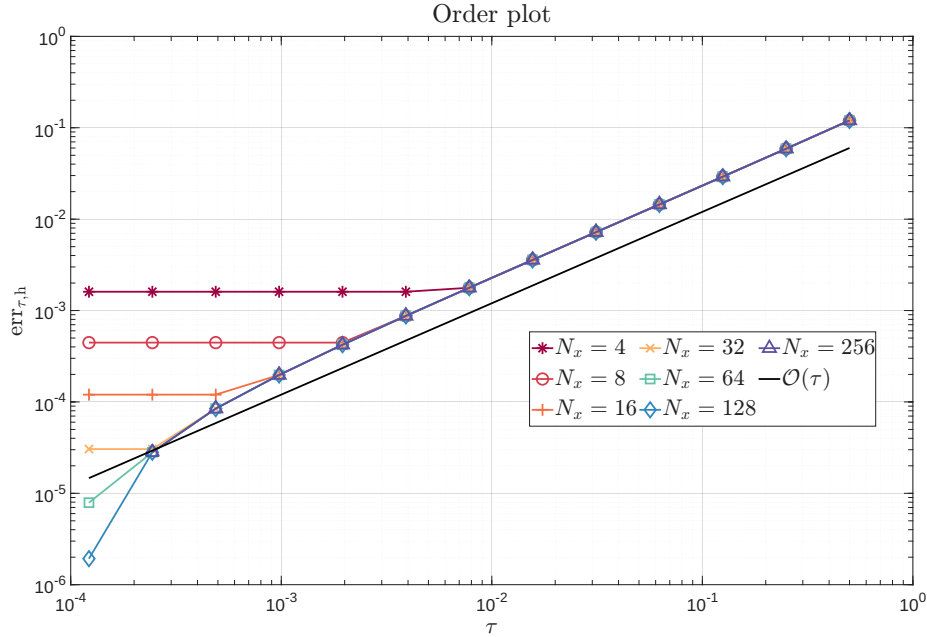


Figure 2: The errors $\text{err}_{\tau,h}$ plotted versus τ for different h . The errors converge like $\mathcal{O}(\tau)$ until stagnating due to the spatial error.

classical order of Lie splitting applied to a standard ODE, and the order of the underlying FEM discretization. Thus they cannot be improved upon, except for possibly removing the logarithmic terms. This behaviour is also observed in practice when the stated assumptions are fulfilled. These are fairly mild, in the sense that they cover all DREs arising from distributed control LQR applications, as long as the initial condition P_0 and the coefficients Q and S are sufficiently regular. Regularity of P_0 is required also for first-order convergence of Lie splitting applied to the DRE without a spatial discretization [17]. Thus the only new assumptions compared to [17] are the ones on Q and S and on the properties of the spatial discretization. The latter is very general and covers many standard FEM discretizations. The former is required to achieve a true $\mathcal{O}(h^2 + |\log h|)$ bound on the spatial error, but can be dropped if a term $h^2(n\tau)^{-1}$ in the error bound is deemed tolerable.

We note that the analysis does not cover DREs arising from boundary control applications. These result in either an unbounded operator B or an unbounded C , or both. In either case, the DRE solution suffers a loss of regularity, which has a direct effect on the speed of convergence that any discretization can achieve. Simply stating such problems properly requires a much more technical framework and as far as we are aware there are no spatial convergence results similar to [19] or our new analysis. A significant amount of preliminary work is therefore required before a rigorous analysis of a full discretization such as the one

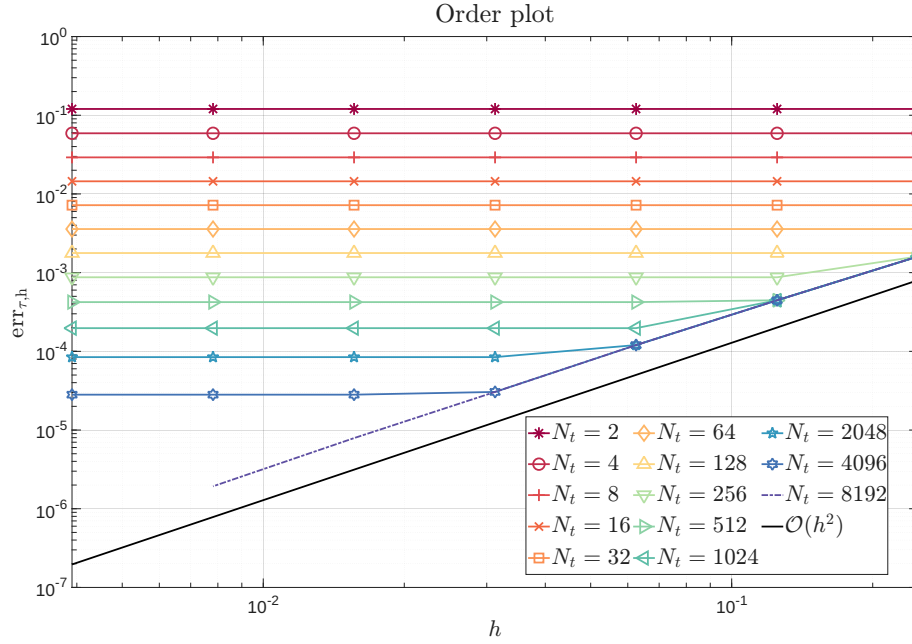


Figure 3: The errors $\text{err}_{\tau,h}$ plotted versus h for different τ . The errors converge like $\mathcal{O}(h^2)$ until stagnating due to the temporal error.

presented here can be attempted. It is our aim to consider this in the future.

Acknowledgments

The authors were partially supported by the Swedish Research Council under grants 2023-03982 (EH, TS, TÅ) and 2023-04862 (EH, TS).

References

- [1] P. Benner and N. Lang. Peer methods for the solution of large-scale differential matrix equations, 2018.
- [2] P. Benner and H. Mena. BDF methods for large-scale differential Riccati equations. In B. De Moor, B. Motmans, J. Willems, P. Van Dooren, and V. Blondel, editors, *Proc. 16th Intl. Symp. Mathematical Theory of Network and Systems, MTNS 2004*, 2004.
- [3] P. Benner and H. Mena. Rosenbrock methods for solving Riccati differential equations. *IEEE Trans. Autom. Control*, 58(11):2950–2956, 2013. ISSN 0018-9286.

- [4] P. Benner and H. Mena. Numerical solution of the infinite-dimensional LQR problem and the associated Riccati differential equations. *J. Numer. Math.*, 26(1):1–20, 2018. doi: 10.1515/jnma-2016-1039.
- [5] 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 edition, 2007. ISBN 978-0-8176-4461-1; 0-8176-4461-X.
- [6] S. C. Brenner and L. R. Scott. *The Mathematical Theory of Finite Element Methods*. Texts in Applied Mathematics. Springer, New York, third edition, 2008. ISBN 978-0-387-75933-3; 978-0-387-75934-0.
- [7] J. Burns and J. Cheung. Optimal convergence rates for Galerkin approximation of operator Riccati equations. *Numer. Methods Partial Differential Eq.*, 38(6):2045–2083, 2022. ISSN 0749-159X. doi: 10.1002/num.22863.
- [8] H. Chen and A. Borzi. Low-rank exponential integrators for stiff differential Riccati equations. *Adv. Comput. Math.*, 51(2):Paper No. 17, 31pp, 2025. ISSN 1019-7168,1572-9044. doi: 10.1007/s10444-025-10228-w.
- [9] J. Cheung. On the approximation of operator-valued Riccati equations in Hilbert spaces. *J. Math. Anal. Appl.*, 547(1):129250, 2025. ISSN 0022-247X. doi: 10.1016/j.jmaa.2025.129250.
- [10] P. G. Ciarlet. *The Finite Element Method for Elliptic Problems*. Classics in Applied Mathematics. SIAM, Philadelphia, PA, 2002. ISBN 978-0-898715-14-9.
- [11] M. Crouzeix. Parabolic evolution problems. <https://perso.univ-rennes1.fr/michel.crouzeix/publis/pabolic.pdf>. Last visited on 2025/11/06.
- [12] R. Curtain and H. Zwart. *Introduction to infinite-dimensional systems theory*, volume 71 of *Texts in Applied Mathematics*. Springer, 2020. ISBN 978-1-0716-0588-2; 978-1-0716-0590-5. doi: 10.1007/978-1-0716-0590-5. A state-space approach.
- [13] E. Emmrich. Discrete versions of Gronwall’s lemma and their application to the numerical analysis of parabolic problems. Technical Report 637, TU Berlin, Inst. Mathematik, 1999.
- [14] D. Gilbarg and N. S. Trudinger. *Elliptic partial differential equations of second order*. Classics in Mathematics. Springer, Berlin Heidelberg, third edition, 2001. ISBN 978-3-540-41160-4; 978-3-642-61798-0. doi: 10.1007/978-3-642-61798-0.
- [15] W. Hackbusch. *Elliptic differential equations*. Springer Series in Computational Mathematics. Springer, Berlin, second edition, 2017. ISBN 978-3-662-54960-5; 978-3-662-54961-2. doi: 10.1007/978-3-662-54961-2.

- [16] E. Hansen and T. Stillfjord. Convergence analysis for splitting of the abstract differential Riccati equation. *SIAM J. Numer. Anal.*, 52(6):3128–3139, 2014. ISSN 0036-1429. doi: 10.1137/130935501.
- [17] E. Hansen, T. Stillfjord, and T. Åberg. Convergence analysis of Lie and Strang splitting for operator-valued differential Riccati equations. *SIAM J. Numer. Anal.*, 64(3):957–988, 2026. ISSN 0036-1429. doi: 10.1137/25M1754406.
- [18] M. Kroller. *Numerische approximation des linear quadratischen optimierungsproblems bei parabolischen differentialgleichungen*. PhD thesis, Technische Universität Graz, 1986. In German.
- [19] M. Kroller and K. Kunisch. Convergence rates for the feedback operators arising in the linear quadratic regulator problem governed by parabolic equations. *SIAM J. Numer. Anal.*, 28(5):1350–1385, 1991. ISSN 0036-1429. doi: 10.1137/0728071.
- [20] S. Larsson and V. Thomée. *Partial differential equations with numerical methods*, volume 45 of *Texts in Applied Mathematics*. Springer, Berlin, 2009. ISBN 978-3-540-88705-8. Paperback reprint of the 2003 edition.
- [21] I. Lasiecka. Convergence estimates for semidiscrete approximations of non-selfadjoint parabolic equations. *SIAM J. Numer. Anal.*, 21(5):894–909, 1984. ISSN 0036-1429. doi: 10.1137/0721058.
- [22] I. Lasiecka and R. Triggiani. *Control theory for partial differential equations: continuous and approximation theories. I*, volume 74 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 2000. ISBN 0-521-43408-4. Abstract parabolic systems.
- [23] D. Li, X. Zhang, and R. Liu. Exponential integrators for large-scale stiff Riccati differential equations. *J. Comput. Appl. Math.*, 389:113360, 17, 2021. ISSN 0377-0427. doi: 10.1016/j.cam.2020.113360.
- [24] A. Målqvist, A. Persson, and T. Stillfjord. Multiscale differential Riccati equations for linear quadratic regulator problems. *SIAM J. Sci. Comput.*, 56(5):3598–3618, 2018. doi: 10.1137/18M1178815.
- [25] H. Mena, A. Ostermann, L.-M. Pfurtscheller, and C. Piazzola. Numerical low-rank approximation of matrix differential equations. *J. Comput. Appl. Math.*, 340:602–614, 2018. doi: 10.1016/j.cam.2018.01.035.
- [26] A. Ostermann, C. Piazzola, and H. Wlach. Convergence of a low-rank Lie-Trotter splitting for stiff matrix differential equations. *SIAM J. Numer. Anal.*, 57(4):1947–1966, 2019. ISSN 0036-1429. doi: 10.1137/18M1177901.
- [27] M. Renardy and R. C. Rogers. *An introduction to partial differential equations*, volume 13 of *Texts in Applied Mathematics*. Springer, second edition, 2004. ISBN 0-387-00444-0.

- [28] I. G. Rosen. Convergence of Galerkin approximations for operator Riccati equations—a nonlinear evolution equation approach. *J. Math. Anal. Appl.*, 155(1):226–248, 1991. ISSN 0022-247X. doi: 10.1016/0022-247X(91)90035-X.
- [29] T. Stillfjord. Low-rank second-order splitting of large-scale differential Riccati equations. *IEEE Trans. Autom. Control*, 60(10):2791–2796, 2015. doi: 10.1109/TAC.2015.2398889.
- [30] T. Stillfjord. Convergence analysis for the exponential Lie splitting scheme applied to the abstract differential Riccati equation. Technical report, Lund University preprints, 2015.
- [31] T. Stillfjord. Adaptive high-order splitting schemes for large-scale differential Riccati equations. *Numer. Algorithms*, 78(4):1129–1151, 2018. ISSN 1017-1398. doi: 10.1007/s11075-017-0416-8.
- [32] T. Stillfjord. Experiment parameters for publication “Convergence analysis of a full discretization of operator-valued differential Riccati equations”. Zenodo, 2026. doi: 10.5281/zenodo.22690855.
- [33] V. Thomée. *Galerkin finite element methods for parabolic problems*, volume 25 of *Springer Series in Computational Mathematics*. Springer, Berlin, 1997. ISBN 3-540-63236-0. doi: 10.1007/978-3-662-03359-3.

A Verification of analyticity

Lemma A.1. *Let $H = L^2(\Omega)$ with $\Omega = (0, 1) \times (0, 1)$ and let $A = \Delta$ with $\mathcal{D}(A) = H_{\mathbb{T}}^2(\Omega)$. Then A generates an analytic semigroup.*

Proof. By, e.g., Theorem 12.31 in [27], A generates an analytic semigroup if there exists a $\omega \in \mathbb{R}$ such that the resolvent set of A contains the half-plane $\{\lambda ; \Re(\lambda) > \omega\}$ and the resolvent $R_\lambda(A) = (A - \lambda I)^{-1}$ satisfies

$$\|R_\lambda(A)\|_{\mathcal{L}(H)} \leq \frac{C}{\lambda - \omega}$$

for some $C > 0$. Here, it is well-known that the spectrum of A is given by $\{\lambda_{j,k} = \lambda_j + \lambda_k, ; i, j = 1, 2, \dots\}$ where $\lambda_j = -j^2\pi^2$ if j is even and $\lambda_j = -(j-1)^2\pi^2$ if j is odd. Thus the spectrum is contained in the left half-plane $\{\lambda ; \Re(\lambda) \leq 0\}$ and the first condition is satisfied with $\omega = 0$. To verify the second condition, we let $R_\lambda(A)u = v$. Then by integration by parts,

$$(u, v) = (Av - \lambda v, v) = -(\nabla v, \nabla v) - \lambda \|v\|^2 \leq -\lambda \|v\|^2,$$

due to the periodic boundary conditions. But by Cauchy-Schwarz inequality, we also have $(u, v) \geq -\|u\|\|v\|$. Hence

$$\|u\| \geq \lambda \|v\|.$$

Now inserting this in $\|R_\lambda(A)\|_{\mathcal{L}(H)} = \sup_{u \neq 0} \frac{\|R_\lambda(A)u\|}{\|u\|}$ proves the assertion. $\square$