Convergence analysis of Lie and Strang splitting for operator-valued differential Riccati equations

Eskil Hansen   Tony Stillfjord   Teodor Åberg

April 25, 2025

Abstract

Differential Riccati equations (DREs) are semilinear matrix- or operator-valued differential equations with quadratic non-linearities. They arise in many different areas, and are particularly important in optimal control of linear quadratic regulators, where they provide the optimal feedback control laws. In the context of control of partial differential equations, these Riccati equations are operator-valued. To approximate their solutions, both spatial and temporal discretizations are needed. While the former have been well analyzed in the literature, there are very few rigorous convergence analyses of time stepping methods applied to DREs, particularly in the infinite-dimensional, operator-valued setting. In view of this, we analyze two numerical time-stepping schemes, the Lie and Strang splitting methods, in such a setting. The analysis relies on the assumption that the uncontrolled system evolves via an operator that generates an analytic semigroup, and that either the initial condition is sufficiently smooth, or the nonlinearity in the DRE is sufficiently smoothing. These assumptions are mild, in the sense that they are not enough to even guarantee continuity in operator-norm of the exact solution to the DRE. However, they imply certain regularity in a pointwise sense, which can be leveraged to prove convergence in operator-norm with the classical orders. The results are illustrated by four numerical experiments, where convergence with the expected order is correlated with the relevant assumptions being fulfilled. The experiments also demonstrate that matrix-valued DREs which arise as spatial discretizations of operator-valued DREs behave similarly, unless the discretization is coarse.

1 Introduction

A differential Riccati equation (DRE) is of the form

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

where the given data A, Q and S and the solution values $P(t)$ are either real matrices or linear operators on a Hilbert space. By A^* we mean either the adjoint of A if it is an operator, or its Hermitian transpose if it is a matrix.

A typical application where DREs arise is the linear quadratic regulator (LQR) problem, where one aims to steer the state vector $x(t)$ to a desirable state by adjusting an input function $u(t)$. The state x satisfies

$$\dot{x} = Ax + Bu,$$
$$y = Ex,$$

with given matrices or operators A, B and E, and the typical goal is to minimize the cost functional

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

where G is a penalization operator for the final state. It is well-known that the optimal feedback control strategy for doing so is $u(t) = -B^*P(T-t)x(t)$, where P satisfies the DRE (1.1) with $Q = E^*E$, $S = BB^*$ and $P_0 = G^*G$.

In the finite-dimensional LQR case, $x(t) \in \mathbb{R}^N$ for some N and the above constitutes control of a system of N ordinary differential equations (ODEs). This leads to a matrix-valued DRE. If $x(t)$ is instead a function in, e.g., $L^2(\Omega)$, where Ω is the spatial domain, we are concerned with control of a partial differential equation (PDE) and the corresponding DRE is operator-valued. Often, a system of ODEs arises as a spatial discretization of a PDE. In this case, when the discretization is refined, the solution to the matrix-valued DRE should tend to the solution of the corresponding operator-valued DRE.

In this paper, we analyze two time-stepping methods applied to operator-valued DREs. As indicated above, there are two main reasons for focusing on operator-valued rather than matrix-valued equations. First, it allows us to ignore the spatial discretization, and thereby decouple the temporal and spatial errors. A full discretization can be acquired by subsequently discretizing the time-discrete system in space. Secondly, given that finer spatial discretizations typically result in systems with less nice properties, the operator-valued DRE also corresponds to the "worst" possible spatial discretization in terms of analysis. Our temporal convergence results will automatically hold for the spatially discretized matrix-valued equations under minimal assumptions, regardless of how fine or coarse these discretizations are.

As our numerical experiments show, the fact that a time-stepping method is convergent for a DRE in $R^{N \times N}$ for any N does not imply that it is convergent in the limit as $N \to \infty$. Basically, the error constants may depend on N in non-obvious ways, or on regularity properties that only become apparent for large N. This means that a large N might require unfeasibly small time steps before any decrease in the temporal errors is observed, unless the problem satisfies certain properties. Our operator-valued analysis exposes these properties.

The time-stepping methods that we are concerned with here are the Lie and Strang splitting schemes. Instead of approximating the solution to the full DRE (1.1) directly, these methods work by combining approximations of the

solutions to the two subproblems $\dot{P} = FP$ and $\dot{P} = GP$, where, informally,

$$FP = A^*P + PA + Q,$$
$$GP = -PSP.$$

We denote the corresponding solution operators by e^{tF} and e^{tG}, respectively. Then with the approximation $P^n \approx P(n\tau)$, the Lie splitting method is given by

$$P^{n+1} = e^{\tau F}e^{\tau G}P^n,$$

and the Strang splitting is defined by

$$P^{n+1} = e^{\tau/2F}e^{\tau G}e^{\tau/2F}P^n.$$

How to properly interpret and define the operators F and G, as well as the solution operators e^{tF} and e^{tG} will be covered in Section 2.

Many other time-stepping methods for (matrix-valued) DRE have been suggested in the literature, e.g. BDF methods [2], Rosenbrock methods [3], Peer methods [1], and exponential integrators [10]. The splitting schemes considered in this paper were introduced in [18], see also [20]. The same methods but with a different decomposition of the right-hand-side were considered in [13, 14]. A common, but surprising, theme in all these works is a lack of rigorous convergence analysis. We are only aware of three exceptions: First, [7] considers an operator-valued setting, but the analyzed method does not fully take advantage of the structure of the DRE and is therefore less efficient than the methods we consider here. Secondly, the methods considered in [14] are similar to ours, but uses a different decomposition of the right-hand-side that typically leads to larger errors for the same step size. Additionally, the analysis is only for the matrix-valued case. Finally, the unpublished manuscript [19] considers the same methods as in this paper in an operator-valued setting. However, this particular setting is too restrictive and the analysis is therefore only applicable to very few problems. The main goal of the present paper is thus to remedy this situation and provide a rigorous convergence analysis for Lie and Strang splitting that is applicable also in an operator-valued setting.

We note that spatial discretizations of operator-valued DREs have been analyzed to a much greater extent than temporal discretizations. Convergence without order is shown in e.g. [17, 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 [8]. The comprehensive book [9] proves convergence with orders also in more general settings, but only for *algebraic* Riccati equations. These results can, however, likely be extended to DREs.

Informally, the main assumptions on (1.1) in this paper is that A generates an analytic semigroup (see, e.g., [15, Chapter 2.5]), and that B and E are bounded operators. This allows for applications such as distributed control. Given this, our main result is that Lie and Strang splitting are convergent with the expected orders 1 and 2, respectively, under certain regularity assumptions. For Lie, we (again informally) require either that the initial condition P_0 is in the domain of

3

$P \mapsto A^*P + PA$, or that S satisfies a similar condition. The former condition is less restrictive than the usual one, where the operator would be applied twice. This is possible due to the smoothing properties of A. In the latter case, P_0 can be arbitrarily non-smooth, but the smoothness assumption on S means that the solution $P(t)$ is nevertheless sufficiently smooth. For Strang splitting, we require both the above assumptions.

A main difficulty here is that the above-mentioned regularity only holds away from $t = 0$. The solution $t \mapsto P(t)$ is not differentiable or even continuous at $t = 0$ in operator-norm, even under the above assumptions. Thus the usual Taylor expansions which could be done in the matrix-valued case are no longer valid. However, the assumptions guarantee that for any x, $t \mapsto P(t)x$ is either continuous or continuously differentiable, and P or $\dot{P}$ is uniformly bounded. This allows us to instead expand around $t = \epsilon > 0$ and by careful manipulation of the rest terms prove that they are uniformly bounded independent of ϵ.

The paper is organized as follows. In Section 2, we discuss the abstract setting and make the above informal statements mathematically rigorous. We also state and prove some preliminary lemmas on existence and smoothness of solutions of the DRE. The convergence analysis of Lie splitting is presented in Section 3, and followed by the analysis of Strang splitting in Section 4. These results are illustrated by three numerical experiments in Section 5, where the different assumptions are either satisfied or not. By repeatedly refining a spatial discretization, we approach the operator-valued case. Finally, we present our conclusions and ideas for future work in Section 6.

2 Setting

Given a Hilbert space X, we use $(\cdot, \cdot)_X$ to denote its inner product, or simply $(\cdot, \cdot)$ when the particular space is clear from the context. The induced operator norm is denoted by $\|\cdot\|_X$ or simply $\|\cdot\|$. Given two general Hilbert spaces X and Y, we denote by $\mathcal{L}(X, Y)$ the space of linear bounded operators from X to Y. If $X = Y$, we abbreviate $\mathcal{L}(X) = \mathcal{L}(X, X)$ If $L : X \to Y$ then $L^* : Y \to X$ denotes the Hilbert space adjoint of L that satisfies $(Lx, y)_Y = (x, L^*y)_X$ for all $x \in X$ and $y \in Y$.

Now let H, U and Y be specific given Hilbert spaces, with H complex. Since the solution $P(t)$ should be symmetric and positive definite, we will consider the Banach space $\Sigma(H) = \{P \in \mathcal{L}(H) \mid P = P^*\}$ with the $\mathcal{L}(H)$-norm and the cone $\Sigma^+(H) = \{P \in \Sigma(H) \mid P \geq 0\}$. We make the following assumption on the operators A, Q and S in (1.1):

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 $A : \mathcal{D}(A) \subset H \to H$ is the generator of an analytic semigroup $t \mapsto \mathrm{e}^{tA} \in \mathcal{L}(H)$.*

Remark 2.1. It follows directly from the assumption that $Q, S \in \Sigma^+(H)$. Further, A is closed and $\mathcal{D}(A)$ is dense in H [15, Corollary 1.2.5].

Remark 2.2. In the context of LQR problems, we assume that the control operator B and the observation operator E are both bounded. This allows for distributed control and observation but typically not boundary control and observation. Extending the work to less regular B and E is likely possible, but at the cost of reduced convergence orders and significantly more technical machinery.

In the matrix-valued case when $P(t)$ and A are in $\mathbb{R}^{N \times N}$, the right-hand side of (1.1) is well defined. In our context, we are interested in operator-valued problems and search for a solution P with $P(t) \in \Sigma^+(H) \subset \Sigma(H)$. Under Assumption 1, the terms Q and $P(t)SP(t)$ are still well defined as operators in $\Sigma(H)$. However, since A is unbounded we need to properly interpret $A^*P(t) + P(t)A$, which is not necessarily in $\Sigma(H)$. We follow [5] and define $\mathcal{A} : \mathcal{D}(\mathcal{A}) \subset \Sigma(H) \to \Sigma(H)$ by

$$(\mathcal{A}(P)x, y) = \varphi_P(x, y), \quad x, y \in \mathcal{D}(A), P \in \mathcal{D}(\mathcal{A}),$$

where

$$\varphi_P(x, y) = (Px, Ay) + (Ax, Py), \quad x, y \in \mathcal{D}(A),$$

and

$$\mathcal{D}(\mathcal{A}) = \{P \in \Sigma(H) \, ; \, \varphi_P \text{ is continuous in } H \times H\}.$$

Essentially this means that $\mathcal{D}(\mathcal{A})$ are those $P \in \Sigma(H)$ for which $A^*P + PA$ can be extended to an operator in $\Sigma(H)$. On $\mathcal{D}(A)$, this extension agrees with the original expression:

Lemma 2.1 (Proposition 3.1 [5]). *If $P \in \mathcal{D}(\mathcal{A})$ then $x \in \mathcal{D}(A)$ implies that $Px \in \mathcal{D}(A^*)$ and*

$$\mathcal{A}Px = A^*Px + PAx.$$

With this in mind, for the rest of the paper we will be considering the equation

$$\dot{P} = \mathcal{A}P + Q - PSP \tag{2.1}$$

rather than (1.1). According to Lemma 2.1, if $P(t) \in \mathcal{D}(\mathcal{A})$ then these equations coincide after multiplication with $x \in \mathcal{D}(A)$.

The operator A is a generator of an analytic semigroup e^{tA}, and this gives rise to an analytic semigroup $\mathrm{e}^{t\mathcal{A}} : \Sigma(H) \to \Sigma(H)$ generated by $\mathcal{A}$, defined by

$$\mathrm{e}^{t\mathcal{A}}P = \mathrm{e}^{tA^*}P\mathrm{e}^{tA}, \quad t \geq 0.$$

As noted in [5, Remark 3.1], it is not strongly continuous, i.e. $\|\mathrm{e}^{t\mathcal{A}}P - P\|_{\mathcal{L}(H)} \not\to 0$ as $t \to 0^+$. However, it satisfies $\|\mathrm{e}^{t\mathcal{A}}Px - Px\| \to 0$ as $t \to 0^+$ for every $x \in H$ and $P \in \Sigma(H)$.

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

Lemma 2.2. *There exist a constant C such that the analytic semigroup e^{tA} generated by the operator A satisfies the inequalities*

$$\|\mathrm{e}^{tA}\|_{\mathcal{L}(H)} \leq C \quad and \quad \|A\mathrm{e}^{tA}\|_{\mathcal{L}(H)} \leq \frac{C}{t}$$

for $t \in (0, T]$. Similarly, the analytic semigroup $\mathrm{e}^{t\mathcal{A}}$ satisfies

$$\|\mathrm{e}^{t\mathcal{A}}P\|_{\mathcal{L}(H)} \leq C\|P\|_{\mathcal{L}(H)} \quad and \quad \|\mathcal{A}\mathrm{e}^{t\mathcal{A}}P\|_{\mathcal{L}(H)} \leq \frac{C}{t}\|P\|_{\mathcal{L}(H)}$$

for $t \in (0, T]$ and $P \in \Sigma(H)$.

Proof. The first two inequalities are standard results for analytic semigroups, see e.g. [16, Lemma 12.32]. They are usually stated with $Ce^{\omega t}$ instead of C, where $\omega \in \mathbb{R}$, but the exponential is clearly bounded by $\max(1, e^{\omega T})$ on the finite time interval $[0, T]$.

Since A is closed (Remark 2.1), it holds that $(A^*)^* = A$ [16, Theorem 8.5.7] and thus $\left(\mathrm{e}^{tA^*}\right)^* = \mathrm{e}^{tA}$ [15, Corollary 1.10.6]. This means that the bound for $\mathrm{e}^{t\mathcal{A}}$ follows directly by applying the bounds of the individual semigroups e^{tA} and e^{tA^*}. We next verify that $\mathrm{e}^{t\mathcal{A}}P$ lies in $\mathcal{D}(\mathcal{A})$. This holds, since for any $t \in (0, T]$ and $x, y \in \mathcal{D}(A)$,

$$
\begin{aligned}
|\varphi_{\mathrm{e}^{t\mathcal{A}}P}(x, y)| &= |(Ax, \mathrm{e}^{tA^*}P\mathrm{e}^{tA}y) + (\mathrm{e}^{tA^*}P\mathrm{e}^{tA}x, Ay)| \\
&= |(\mathrm{e}^{tA}Ax, P\mathrm{e}^{tA}y) + (P\mathrm{e}^{tA}x, \mathrm{e}^{tA}Ay)| \\
&= |(A\mathrm{e}^{tA}x, P\mathrm{e}^{tA}y) + (P\mathrm{e}^{tA}x, A\mathrm{e}^{tA}y)| \\
&\leq \|A\mathrm{e}^{tA}\|_{\mathcal{L}(H)}\|x\|\|P\|_{\mathcal{L}(H)}\|\mathrm{e}^{tA}\|_{\mathcal{L}(H)}\|y\| \\
&\quad + \|A\mathrm{e}^{tA}\|_{\mathcal{L}(H)}\|x\|\|P\|_{\mathcal{L}(H)}\|\mathrm{e}^{tA}\|_{\mathcal{L}(H)}\|y\| \\
&\leq \frac{C}{t}\|P\|_{\mathcal{L}(H)}\|x\|\|y\|.
\end{aligned}
$$

As noted in Remark 2.1, $\mathcal{D}(A)$ is dense in H, and the bound thus holds for all $x, y \in H$. But, by definition,

$$\|\mathcal{A}\mathrm{e}^{t\mathcal{A}}P\|_{\mathcal{L}(H)} = \sup_{x \in H} \sup_{y \in H} \frac{|(\mathcal{A}(P)x, y)|}{\|x\|\|y\|} = \sup_{x \in H} \sup_{y \in H} \frac{|\varphi_{\mathrm{e}^{t\mathcal{A}}P}(x, y)|}{\|x\|\|y\|},$$

so the above inequality finishes the proof. $\qquad\square$

Our basic assumptions on the operators agree with those in [5], and imply the existence of a unique solution to the problem:

Lemma 2.3 ([5, Theorem IV-1.3.1])**.** *Under Assumption 1, there is a (classical) solution P to (2.1) on $t \in [0, T]$. That is, $P(0) = P_0$ and for any $t \in (0, T)$, $P(t)$ is differentiable with $P(t) \in \mathcal{D}(\mathcal{A}) \cap \Sigma^+(H)$ and $\dot{P}(t) = \mathcal{A}P(t) + Q - P(t)SP(t)$. Further, for any $\varepsilon > 0$, $P \in C^\infty\big([\varepsilon, T], \Sigma(H)\big)$.*

Here,

$$C^\infty\big([\varepsilon,T],\Sigma(H)\big) = \begin{cases} t \mapsto P(t) \in \Sigma(H) \text{ is infinitely differentiable in } \mathcal{L}(H) \\[2mm] \text{for } t \in [\varepsilon,T] \text{ and } \sup_{t\in(\varepsilon,T)} \|\frac{\mathrm{d}^k}{\mathrm{d}t^k}P(t)\|_{\mathcal{L}(H)} < \infty,\ k = 0,1,\dots \end{cases}.$$

Thus a solution exists regardless of how smooth P_0 is, and it is infinitely smooth away from $t = 0$. However, the solution will typically not be continuous at $t = 0^+$ in the $\mathcal{L}(H)$-norm. We can only guarantee that P is strongly continuous. We therefore introduce the function space

$$C_p\big([0,T],\Sigma(H)\big) = \begin{cases} t \mapsto P(t) \in \Sigma(H) \mid t \mapsto P(t)x \text{ is continuous in } H \\[2mm] \text{for any } x \in H \end{cases},$$

equipped with the norm

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

As noted in [5, Section IV-1.2.1], $\|P\|_\infty$ is finite for any $P \in C_p\big([0,T],\Sigma(H)\big)$ by the uniform boundedness theorem.

We also consider the spaces $C_p^k\big([0,T],\Sigma(H)\big)$ which consist of functions $t \mapsto P(t) \in \Sigma(H)$ such that $t \mapsto P(t)x$ is k times differentiable for any $x \in H$ and where the k-th derivatives belongs to $C_p\big([0,T],\Sigma(H)\big)$. With these definitions, we have the following regularity near $t = 0$:

Lemma 2.4 ([5, Proposition IV-1.3.2 and Theorem IV-1.3.1]). *Under Assumption 1 the solution P guaranteed to exist by Lemma 2.3 satisfies $P \in C_p\big([0,T],\Sigma(H)\big)$. If in addition $P_0 \in \mathcal{D}(\mathcal{A})$, then $P \in C_p^1\big([0,T],\Sigma(H)\big)$ and for any $t \in [0,T]$, $\mathcal{A}P \in C_p\big([0,T],\Sigma(H)\big)$.*

Remark 2.3. In [5], the function spaces $C_s([0,T],\Sigma(H))$ and $C_u([0,T],\Sigma(H))$ are considered, and the above results are stated as $P \in C_s([0,T],\Sigma(H))$. However, these two spaces contain the same sets of functions, and the results can thus equivalently be stated for $P \in C_u([0,T],\Sigma(H))$. We prefer to use this version here because $C_u([0,T],\Sigma(H))$ is a Banach space, in contrast to $C_s([0,T],\Sigma(H))$ which is equipped with a weaker topology. Our space $C_p\big([0,T],\Sigma(H)\big)$ is the same as $C_u([0,T],\Sigma(H))$ but has been renamed in an attempt to minimize confusion. The subscript p is meant to be read as "pointwise" (for points $x \in H$).

We also have positive semi-definiteness:

Lemma 2.5 ([5, Proposition IV-1.3.3 and Theorem IV-1.2.1]). *For each $t \in [0,T]$, the solution P to (2.1) guaranteed by Lemma 2.4 satisfies $P(t) \in \Sigma^+(H)$.*

We now define the operators $F : \mathcal{D}(\mathcal{A}) \to \mathcal{L}(H)$ and $G : \mathcal{L}(H) \to \mathcal{L}(H)$ by

$$FP = \mathcal{A}P + Q, \tag{2.2}$$
$$GP = -PSP, \tag{2.3}$$

and consider the two subproblems of (2.1) that arise when we either omit the GP part or the FP part of the equation:

Corollary 2.6. *Under Assumption 1 there exist unique solutions to the two subproblems*

$$\dot{P} = FP, \quad P(0) = P_0 \tag{2.4}$$

and

$$\dot{P} = GP, \quad P(0) = P_0, \tag{2.5}$$

which satisfy the stated regularity properties in Lemma 2.3 and Lemma 2.4. For the second subproblem, the conclusions of Lemma 2.4 hold also without $P_0 \in \mathcal{D}(\mathcal{A})$.

Proof. The subproblems can be seen as instances of (2.1), where we either take $S = 0$ or we take $A = Q = 0$. These choices certainly fulfil the respective parts of Assumption 1, so we can apply the previous lemmas to establish the existence and regularity properties. The final assertion follows simply by the fact that $\mathcal{D}(\mathcal{A}) = \Sigma(H)$ if $A = 0$. $\square$

We will frequently use the notation $t \mapsto \mathrm{e}^{tF}P_0$ and $t \mapsto \mathrm{e}^{tG}P_0$ to refer to the solutions to the subproblems (2.4) and (2.5). In the case of (2.4) we also have the representation

$$\mathrm{e}^{tF}P_0 = \mathrm{e}^{t\mathcal{A}}P_0 + \int_0^t \mathrm{e}^{(t-s)\mathcal{A}}Q \,\mathrm{d}s,$$

and the lack of A in (2.5) allows us to say more than Corollary 2.6:

Lemma 2.7. *Under Assumption 1, the solution to the problem (2.5) is given by $\mathrm{e}^{tG}P_0 = (I + tP_0S)^{-1}P_0$, where*

$$\|t \mapsto \mathrm{e}^{tG}P_0\|_\infty \leq \big(1 + T\rho\|S\|_{\mathcal{L}(H)}\big)\rho$$

for every P_0 with $\|P_0\| \leq \rho$. Additionally, the derivatives of $t \mapsto \mathrm{e}^{tG}P_0$ are bounded by

$$\left\|t \mapsto \frac{\mathrm{d}^j}{\mathrm{d}t^j}\mathrm{e}^{tG}P_0\right\|_\infty \leq j!\|S\|_{\mathcal{L}(H)}^j\big(1 + T\rho\|S\|_{\mathcal{L}(H)}\big)^{j+1}\rho^{j+1}$$

for $j = 1, 2, \ldots$.

Proof. We first note that for any $P_1, P_2 \in \Sigma^+(H)$, we have

$$\|(I + P_1P_2)^{-1}\|_{\mathcal{L}(H)} \leq 1 + \|P_1\|_{\mathcal{L}(H)}\|P_2\|_{\mathcal{L}(H)}$$

see e.g. [9, Lemma 2A.1]. Thus the function $P(t)$ given by

$$P(t) = (I + tP_0S)^{-1}P_0$$

is well defined for all $t \geq 0$, and $\|P(t)\|_{\mathcal{L}(H)} \leq \big(1 + t\|P_0\|_{\mathcal{L}(H)}\|S\|_{\mathcal{L}(H)}\big)\|P_0\|_{\mathcal{L}(H)}$. As a consequence, the first inequality is satisfied. To prove that P solves the

8

problem (2.5), we first note that

$$\|(I + (t+\tau)P_0 S)^{-1}P_0 - (I + tP_0 S)^{-1}P_0\|_{\mathcal{L}(H)} =$$
$$= \left\|\left(I + (t+\tau)P_0 S\right)^{-1}\left[(I + tP_0 S) - \left(I + (t+\tau)P_0 S\right)\right](I + tP_0 S)^{-1}P_0\right\|_{\mathcal{L}(H)}$$
$$= \left\|-\tau\left(I + (t+\tau)P_0 S\right)^{-1}P_0 S(I + tP_0 S)^{-1}P_0\right\|_{\mathcal{L}(H)}$$
$$\leq \tau\|(I + (t+\tau)P_0 S)^{-1}\|_{\mathcal{L}(H)}\|P_0 S\|_{\mathcal{L}(H)}\|(I + tP_0 S)^{-1}\|_{\mathcal{L}(H)}\|P_0 S\|_{\mathcal{L}(H)}$$
$$\leq \tau\left(1 + (t+\tau)\|P_0\|_{\mathcal{L}(H)}\|S\|_{\mathcal{L}(H)}\right)\left(1 + t\|P_0\|_{\mathcal{L}(H)}\|S\|_{\mathcal{L}(H)}\right)\|P_0\|^2_{\mathcal{L}(H)}$$
$$\leq \tau\, C(t, \|P_0\|_{\mathcal{L}(H)}, \|S\|_{\mathcal{L}(H)}),$$

and $t \mapsto P(t)$ is therefore continuous in $\mathcal{L}(H)$. By the same construction we obtain that

$$\lim_{\tau \to 0}\|(P(t+\tau) - P(t))/\tau + P(t)SP(t)\|_{\mathcal{L}(H)} = 0.$$

The function $t \mapsto P(t)$ is thus continuously differentiable and satisfies the equation, i.e., $P(t) = e^{tG}P_0$. The bound for the higher derivatives follows by induction and the chain rule, since e.g.

$$\frac{\mathrm{d}^2}{\mathrm{d}s^2}e^{sG}P_0 = \frac{\mathrm{d}}{\mathrm{d}s}\left(-e^{sG}P_0 S e^{sG}P_0\right)$$
$$= -\left(\frac{\mathrm{d}}{\mathrm{d}s}e^{sG}P_0\right)S e^{sG}P_0 - e^{sG}P_0 S\left(\frac{\mathrm{d}}{\mathrm{d}s}e^{sG}P_0\right).$$

$\square$

We also have the following explicit bounds on the function G and a similar function that appears in the analysis of the Strang splitting method.

Lemma 2.8. *The functions* $G : \mathcal{L}(H) \to \mathcal{L}(H)$ *and* $M : \mathcal{L}(H) \to \mathcal{L}(H)$ *given by*

$$GP = -PSP \quad and \quad MP = PSPSP$$

are both locally Lipschitz continuous, and

$$\|GP_1 - GP_2\|_{\mathcal{L}(H)} \leq 2\rho\|S\|_{\mathcal{L}(H)}\|P_1 - P_2\|_{\mathcal{L}(H)} \tag{2.6}$$
$$\|MP_1 - MP_2\|_{\mathcal{L}(H)} \leq 3\rho^2\|S\|^2_{\mathcal{L}(H)}\|P_1 - P_2\|_{\mathcal{L}(H)} \tag{2.7}$$

for all P_1, P_2 *such that* $\|P_i\|_{\mathcal{L}(H)} \leq \rho$.

Proof. Both claims follow from straightforward calculations. First,

$$\|GP_1 - GP_2\|_{\mathcal{L}(H)}$$
$$\leq \|P_1 SP_1 - P_1 SP_2 + P_1 SP_2 - P_2 SP_2\|_{\mathcal{L}(H)}$$
$$\leq \|P_1 S\|_{\mathcal{L}(H)}\|P_1 - P_2\|_{\mathcal{L}(H)} + \|P_1 - P_2\|_{\mathcal{L}(H)}\|SP_2\|_{\mathcal{L}(H)}$$
$$\leq 2\rho\|S\|_{\mathcal{L}(H)}\|P_1 - P_2\|_{\mathcal{L}(H)},$$

which proves the first bound, and

$$\begin{aligned}
\|MP_1 - &MP_2\|_{\mathcal{L}(H)} \\
&\leq \|P_1 S P_1 S P_1 - P_1 S P_2 S P_2 + P_1 S P_2 S P_2 - P_2 S P_2 S P_2\|_{\mathcal{L}(H)} \\
&\leq \|P_1 S\|_{\mathcal{L}(H)} \|G P_1 - G P_2\|_{\mathcal{L}(H)} \\
&\quad + \|P_1 - P_2\|_{\mathcal{L}(H)} \|S P_2 S P_2\|_{\mathcal{L}(H)} \\
&\leq 3\rho^2 \|S\|_{\mathcal{L}(H)}^2 \|P_1 - P_2\|_{\mathcal{L}(H)},
\end{aligned}$$

which proves the second. $\qquad\square$

Finally, the solution operators for both subproblems, and thus also the splitting schemes, are invariant on $\Sigma^+(H)$:

Lemma 2.9. *Let Assumption 1 be satisfied and let $P \in \Sigma^+(H)$. Then for any $t \geq 0$, also $\mathrm{e}^{tF}P \in \Sigma^+(H)$ and $\mathrm{e}^{tG}P \in \Sigma^+(H)$.*

Proof. Since

$$(\mathrm{e}^{t\mathcal{A}}Px, x) = (\mathrm{e}^{tA^*}P\mathrm{e}^{tA}x, x) = (P\mathrm{e}^{tA}x, \mathrm{e}^{tA}x) \geq 0$$

and $(\mathrm{e}^{tA^*}P\mathrm{e}^{tA})^* = \mathrm{e}^{tA^*}P\mathrm{e}^{tA}$ it follows that $\mathrm{e}^{t\mathcal{A}}P \in \Sigma^+(H)$. Since $Q \in \Sigma^+(H)$, it then follows by the variation-of-constants formula that $\mathrm{e}^{tF}P \in \Sigma^+(H)$.

For the nonlinear subproblem, we first note that since $P \in \Sigma^+(H)$, there exists an operator $P^{1/2} \in \Sigma^+(H)$ such that $P = P^{1/2}P^{1/2}$. By [9, Lemma 2A.1], we have

$$(I + tPS)^{-1} = I - tP^{1/2}(I + tP^{1/2}SP^{1/2})^{-1}P^{1/2}S.$$

Since $\mathrm{e}^{tG}P = (I + tPS)^{-1}P$ by Lemma 2.7, this means that

$$\begin{aligned}
\mathrm{e}^{tG}P &= P^{1/2}P^{1/2} - P^{1/2}\Big((I + tP^{1/2}SP^{1/2})^{-1}tP^{1/2}SP^{1/2}\Big)P^{1/2} \\
&= P^{1/2}(I + tP^{1/2}SP^{1/2})^{-1}P^{1/2}.
\end{aligned}$$

This expression is self-adjoint and positive semi-definite if $(I + tP^{1/2}SP^{1/2})^{-1}$ is. But if $x \in H$ and $y = (I + tP^{1/2}SP^{1/2})^{-1}x$ then

$$\begin{aligned}
((I + tP^{1/2}SP^{1/2})^{-1}x, x) &= (y, y + tP^{1/2}SP^{1/2}y) \\
&= (y, y) + t(P^{1/2}y, SP^{1/2}y) \geq 0
\end{aligned}$$

as S is positive semi-definite. Thus $\mathrm{e}^{tG}P \in \Sigma^+(H)$. $\qquad\square$

3 Convergence analysis for Lie splitting

We now turn to the convergence analysis and consider first Lie splitting scheme for (2.1). We consider an equidistant grid $t_n = n\tau$ with time step $\tau = T/N$

and approximate $P^n \approx P(t_n)$. In the notation of Section 2, the method is then defined by $P^{n+1} = \mathcal{L}_\tau P^n$, where the time-stepping operator $\mathcal{L}_\tau$ is given by

$$\mathcal{L}_\tau = \mathrm{e}^{\tau F} \mathrm{e}^{\tau G}.$$

We start by expressing one step of the method in a more useful way. By expanding the nonlinear flow in a first-order Taylor expansion, we get for any $P \in \Sigma^+(H)$ that

$$\mathcal{L}_\tau P = \mathrm{e}^{\tau \mathcal{A}} P + \int_0^\tau \mathrm{e}^{(\tau-s)\mathcal{A}} Q \, \mathrm{d}s + \tau \mathrm{e}^{\tau \mathcal{A}} GP + \int_0^\tau (\tau - s) \mathrm{e}^{\tau \mathcal{A}} \frac{\mathrm{d}^2}{\mathrm{d}s^2} \mathrm{e}^{sG} P \, \mathrm{d}s \tag{3.1}$$
$$=: L_0(P) + L_1 + L_2(P) + L_3(P).$$

Next, we similarly reformulate the exact solution. Since Lemma 2.3 guarantees $P \in C^\infty\big([\varepsilon, T], \Sigma(H)\big)$, we can take any $\varepsilon > 0$ and do the following Taylor expansion :

$$P(\tau)x = P(\tau - \varepsilon + \varepsilon)$$
$$= \mathrm{e}^{(\tau-\varepsilon)\mathcal{A}} P(\varepsilon) + \int_\varepsilon^\tau \mathrm{e}^{(\tau-s)\mathcal{A}} \big(Q + GP(s)\big) \, \mathrm{d}s. \tag{3.2}$$

According to Lemma 2.3, the operator $\mathrm{e}^{(\tau-s)\mathcal{A}}$ maps any initial condition into a $C^\infty\big([0, \tau - \varepsilon], \Sigma(H)\big)$-function. The composition $s \mapsto \mathrm{e}^{(\tau-s)\mathcal{A}} GP(s)$ is therefore in $C^\infty\big([\varepsilon, \tau - \varepsilon], \Sigma(H)\big)$, since $s \mapsto GP(s) \in C^\infty\big([\varepsilon, T], \Sigma(H)\big)$ like P itself. We can therefore expand $\mathrm{e}^{(\tau-s)\mathcal{A}} GP(s)$ too, which yields

$$P(\tau) = \mathrm{e}^{(\tau-\varepsilon)\mathcal{A}} P(\varepsilon) + \int_\varepsilon^\tau \mathrm{e}^{(\tau-s)\mathcal{A}} Q \, \mathrm{d}s + (\tau - \varepsilon) \mathrm{e}^{(\tau-\varepsilon)\mathcal{A}} GP(\varepsilon)$$
$$+ \int_\varepsilon^{\tau-\varepsilon} \int_\varepsilon^s \frac{\mathrm{d}}{\mathrm{d}r} \Big(\mathrm{e}^{(\tau-r)\mathcal{A}} GP(r) \Big) \, \mathrm{d}r \mathrm{d}s + \int_{\tau-\varepsilon}^\tau \mathrm{e}^{(\tau-s)\mathcal{A}} GP(s) \, \mathrm{d}s$$
$$=: I_0(\varepsilon, P_0) + I_1(\varepsilon, P_0) + I_2(\varepsilon, P_0) + I_3(\varepsilon, P_0) + I_4(\varepsilon, P_0).$$

Now let $x \in H$. Since $t \mapsto P(t)x$ is continuous, the first three terms satisfy

$$\lim_{\varepsilon \to 0^+} I_j(\varepsilon, P_0)x = L_j(P_0)x \quad \text{for } j = 0, 1, 2.$$

That is, the first three terms agree with the corresponding terms in the expansion of the scheme given in (3.1). Further,

$$\|s \mapsto \mathrm{e}^{(\tau-s)\mathcal{A}} GP(s)\|_\infty \leq C \|P\|_\infty^2,$$

by Lemma 2.2, so that

$$\lim_{\varepsilon \to 0^+} \|I_4(\varepsilon, P_0)\|_{\mathcal{L}(H)} = 0.$$

As a consequence, also $\lim_{\varepsilon \to 0^+} I_3(\varepsilon, P_0)x$ exists and equals $R(0, x)$, where

$$R(t, x) = \int_0^\tau \int_0^s \frac{\mathrm{d}}{\mathrm{d}r} \Big(\mathrm{e}^{(\tau-r)\mathcal{A}} GP(t+r)x \Big) \, \mathrm{d}r \mathrm{d}s$$

11

for $t \in [0, T - \tau]$. We can do the same expansions around $P(k\tau)$ instead of P_0 and thus have that

$$P((k+1)\tau)x = (L_0 + L_1 + L_2)(P(k\tau))x + R(k\tau, x), \qquad (3.3)$$

Ideally, we would now argue that $\|R(t, x) - L_3(P(k\tau))x\| \leq C\tau^2\|x\|$, which would give us a bound for the local error of the method. However, this does not necessarily hold under only Assumption 1. Thus, we now introduce the following extra assumptions:

Assumption 2. *The initial condition $P_0 \in \mathcal{D}(\mathcal{A})$.*

Assumption 3. *The operators S and AS both belong to $\mathcal{L}(H)$.*

Either Assumption 2 or Assumption 3 will lead to first-order convergence, except for a $|\log \tau|$-factor in the former case. This factor arises from using the analyticity properties of e^{tA} to essentially ignore one application of $\mathcal{A}$, which allows us to only assume $P_0 \in \mathcal{D}(\mathcal{A})$ rather than the usual $P_0 \in \mathcal{D}(\mathcal{A}^2)$. If the latter does hold, the proof can be modified to remove the $|\log \tau|$-factor.

3.1 Lie splitting under Assumption 2

Our approach in the case of Assumption 2 is similar to the one used in [14] for the matrix-valued case. We note first that since the solution P to (2.1) is in $C_p^1\big([0, T], \Sigma(H)\big)$ by Lemma 2.4, there exists a $\gamma > 0$ such that

$$\max\{\|P\|_\infty, \|\dot{P}\|_\infty\} \leq \gamma.$$

We start with the following preliminary result:

Lemma 3.1. *Let Assumptions 1 and 2 be fulfilled and let P be the solution to (2.1). Then the equality*

$$R(t, x) = \int_0^\tau \int_0^s \lambda_1(t, r)x \,\mathrm{d}r\mathrm{d}s + \int_0^\tau \int_0^s \lambda_2(t, r)x \,\mathrm{d}r\mathrm{d}s$$

of, improper, integrals is well defined for all $t \in [0, T - \tau]$ and $x \in H$. Here,

$$\lambda_1(t, r) = \mathcal{A}\mathrm{e}^{(\tau-r)\mathcal{A}}\big(P(t+r)SP(t+r)\big) \quad and$$
$$\lambda_2(t, r) = -\mathrm{e}^{(\tau-r)\mathcal{A}}\big(\dot{P}(t+r)SP(t+r) + P(t+r)S\dot{P}(t+r)\big).$$

Proof. Lemma 2.3 and the chain rule gives that

$$\frac{\mathrm{d}}{\mathrm{d}r}\Big(\mathrm{e}^{(\tau-r)\mathcal{A}}GP(t+r)\Big) = \lambda_1(t, r) + \lambda_2(t, r)$$

for $t+r > 0$ with $0 \leq r < \tau$, where all three functions are in $C^\infty\big([\varepsilon, \tau-\varepsilon], \Sigma(H)\big)$ with respect to r. Since $P \in C_p^1\big([0, T], \Sigma(H)\big)$, one has the bound

$$\sup_{r \in (0, \tau)} \|\lambda_2(t, r)\|_{\mathcal{L}(H)} \leq C\|P\|_\infty\|\dot{P}\|_\infty\|S\|_{\mathcal{L}(H)} \leq C(\gamma) \qquad (3.4)$$

for all $t \in [0, T - \tau]$. That is, $r \mapsto \lambda_2(t, r)$ is uniformly bounded and the integral $\int_0^\tau \int_0^s \lambda_2(t, r)x \,\mathrm{d}r\mathrm{d}s$ is therefore well defined. Furthermore,

$$\lim_{\varepsilon \to 0^+} \int_\varepsilon^{\tau - \varepsilon} \int_\varepsilon^s \lambda_1(t, r)x \,\mathrm{d}r\mathrm{d}s = R(t, x) - \int_0^\tau \int_0^s \lambda_2(t, r)x \,\mathrm{d}r\mathrm{d}s.$$

Hence, the integral $\int_0^\tau \int_0^s \lambda_1(t, r)x \,\mathrm{d}r\mathrm{d}s$ is also well defined and that the sought after equality holds. $\qquad\square$

We can now prove consistency for the Lie scheme:

Lemma 3.2. *Let Assumptions 1 and 2 be fulfilled and let P be the solution to (2.1). If $\|\mathcal{L}_\tau^k P_0\|_{\mathcal{L}(H)} \leq \gamma + 1$ for all $k = 0, \dots, n$, then*

$$\|P((n + 1)\tau) - \mathcal{L}_\tau^{n+1} P_0\|_{\mathcal{L}(H)} \leq D\tau(1 + |\log \tau|),$$

where $D = D(\gamma)$.

Proof. First observe that the hypothesis implies that the first $n + 1$ errors, i.e.

$$E_k = P(k\tau) - \mathcal{L}_\tau^k P_0, \quad \text{with } k = 0, \dots, n,$$

are all elements in $\mathcal{L}(H)$. By (3.3) and (3.1) it follows that

$$\begin{aligned}
E_{n+1}x &= P((n + 1)\tau)x - \mathcal{L}_\tau^{n+1} P_0 x \\
&= L_0(P(n\tau) - \mathcal{L}_\tau^n P_0)x + (L_2(P(n\tau)) - L_2(\mathcal{L}_\tau^n P_0))x \\
&\quad + R(n\tau, x) - L_3(\mathcal{L}_\tau^n P_0)x
\end{aligned}$$

Noting that the first term is $\mathrm{e}^{\tau\mathcal{A}}E_n x$, we can repeatedly apply this equality and obtain the error representation

$$\begin{aligned}
E_{n+1}x = L_0(E_0)x &+ \sum_{k=1}^n \mathrm{e}^{(n-k)\tau\mathcal{A}}(L_2(P(k\tau)) - L_2(\mathcal{L}_\tau^k P_0))x \\
&+ \sum_{k=0}^n \mathrm{e}^{(n-k)\tau\mathcal{A}}(R(k\tau, x) - L_3(\mathcal{L}_\tau^k P_0))x.
\end{aligned}$$

Since $E_0 = 0$, this representation gives the bound

$$\begin{aligned}
\|E_{n+1}x\| \leq C\sum_{k=1}^n \|(L_2(P(k\tau)) - L_2(\mathcal{L}_\tau^k P_0))x\| &+ C\sum_{k=0}^n \|L_3(\mathcal{L}_\tau^k P_0)x\| \\
+ \left\|\sum_{k=0}^n \mathrm{e}^{\tau(n-k)\mathcal{A}}R(k\tau, x)\right\| &=: K_1 + K_2 + K_3.
\end{aligned} \tag{3.5}$$

As $L_2(P) = \tau\mathrm{e}^{\tau\mathcal{A}}GP$, the hypothesis and Lemma 2.8 with $\rho = \gamma + 1$ yield the bound

$$K_1 \leq C(\gamma)\tau \sum_{k=1}^n \|E_k\|_{\mathcal{L}(H)}\|x\|.$$

13

Furthermore, Lemma 2.7 with $\rho = \gamma + 1$, the trivial bound $k\tau \leq T$, and the hypothesis give

$$K_2 \leq C(\gamma)\,\tau\|x\|.$$

In order to bound K_3, we recall Lemma 3.1 and obtain

$$K_3 \leq K_{31} + K_{32},$$

where

$$K_{3j} = \left\|\sum_{k=0}^{n} \mathrm{e}^{\tau(n-k)\mathcal{A}} \int_0^\tau \int_0^s \lambda_j(k\tau, r)x\,\mathrm{d}r\mathrm{d}s\right\|, \quad j = 1, 2,$$

with λ_j given in Lemma 3.1. By (3.4) one has the bound

$$K_{32} \leq C \sum_{k=0}^{n} \left\|\int_0^\tau \int_0^s \lambda_2(k\tau, r)x\,\mathrm{d}r\mathrm{d}s\right\| \leq C(\gamma)\tau\|x\|.$$

To bound K_{31}, we note that

$$\|\lambda_1(t, r)x\| \leq \frac{C}{\tau - r}\|P\|_\infty^2\|S\|_{\mathcal{L}(H)}\|x\|.$$

Directly using the equality $\int_0^\tau \int_0^s \frac{1}{\tau - r}\,\mathrm{d}r\mathrm{d}s = \tau$ on each of the summands does not result in a good enough bound. We only do this for the last summand with $k = n$, i.e.

$$\left\|\int_0^\tau \int_0^s \lambda_1(n\tau, r)x\,\mathrm{d}r\mathrm{d}s\right\| \leq C(\gamma)\tau\|x\|.$$

For the remaining summands, we move the outer exponential into the integrals and interchange it with the one in the integrand by using the relation $\mathcal{A}\mathrm{e}^{t\mathcal{A}} = \mathrm{e}^{t\mathcal{A}}\mathcal{A}$. Lemma 2.2 then gives the bound

$$\left\|\sum_{k=0}^{n-1} \mathrm{e}^{\tau(n-k)\mathcal{A}} \int_0^\tau \int_0^s \lambda_1(k\tau, r)x\,\mathrm{d}r\mathrm{d}s\right\|$$

$$\leq \sum_{k=0}^{n-1} \int_0^\tau \int_0^s \left\|\mathrm{e}^{(\tau-r)\mathcal{A}}\mathcal{A}\mathrm{e}^{\tau(n-k)\mathcal{A}}GP(k\tau + r)x\right\|\,\mathrm{d}r\mathrm{d}s$$

$$\leq C \sum_{k=0}^{n-1} \int_0^\tau \int_0^s \left\|\mathcal{A}\mathrm{e}^{\tau(n-k)\mathcal{A}}GP(k\tau + r)x\right\|\,\mathrm{d}r\mathrm{d}s$$

$$\leq C \sum_{k=0}^{n-1} \int_0^\tau \int_0^s \frac{1}{(n-k)\tau}\|GP(k\tau + r)x\|\,\mathrm{d}r\mathrm{d}s$$

$$\leq C\|GP\|_\infty \tau^2 \sum_{k=0}^{n-1} \frac{1}{(n-k)\tau}\|x\|$$

$$\leq C(\gamma)\,\tau\big(1 + \log n\big)\|x\|.$$

In total,

$$\|E_{n+1}\|_{\mathcal{L}(H)} \leq \sup_{\|x\|=1} K_1 + K_2 + K_3$$

$$\leq C(\gamma)\big(\tau \sum_{k=0}^{n} \|E_k\|_{\mathcal{L}(H)} + \tau(1 + |\log \tau|)\big).$$

The sought after error bound then follows by a discrete Grönwall inequality, see e.g. [6, Theorem 4.1]. □

Remark 3.1. We note that the condition $\|\mathcal{L}_\tau^k P_0\|_{\mathcal{L}(H)} \leq \gamma + 1$ is an arbitrary choice which could be replaced by $\|\mathcal{L}_\tau^k P_0\|_{\mathcal{L}(H)} \leq \gamma + \rho$ for any $\rho > 0$ in order to tune the error constants.

Theorem 3.3. *Let Assumptions 1 and 2 be fulfilled and let P be the solution to (2.1). Then*

$$\|P(n\tau) - \mathcal{L}_\tau^n P_0\|_{\mathcal{L}(H)} \leq D\tau(1 + |\log \tau|) \quad \text{for all } \tau \leq \tau^*,$$

where the constants D, τ^* only depend on the solution $(\|P\|_\infty, \|\dot{P}\|_\infty)$ and the problem data $(T, \|S\|_{\mathcal{L}(H)}, \|Q\|_{\mathcal{L}(H)})$.

Proof. With the notation from Lemma 3.2, there exists an τ^* such that

$$D\tau(1 + |\log \tau|) \leq 1 \quad \text{for all } \tau \leq \tau^*.$$

Let $\tau \leq \tau^*$ and assume that $\|E_k\|_{\mathcal{L}(H)} \leq D\tau(1 + |\log \tau|)$ for $k = 0, \ldots, n$. Then

$$\|\mathcal{L}_\tau^k P_0\|_{\mathcal{L}(H)} \leq \|P(k\tau)\|_{\mathcal{L}(H)} + \|E_k\|_{\mathcal{L}(H)} \leq \gamma + 1$$

and Lemma 3.2 implies that $\|E_{n+1}\|_{\mathcal{L}(H)} \leq D\tau(1 + |\log \tau|)$. As $\|E_0\|_{\mathcal{L}(H)} = 0$, the first-order convergence follows by induction. □

3.2 Lie splitting under Assumption 3

We once more consider the Lie splitting scheme, but with Assumption 3 rather than Assumption 2. In this case, the solution P to (2.1) guaranteed by Lemma 2.4 is only in $C_p\big([0,T], \Sigma(H)\big)$ rather than $C_p^1\big([0,T], \Sigma(H)\big)$. We therefore set choose a $\gamma > 0$ such that

$$\|P\|_\infty \leq \gamma$$

in this section. We start the analysis by showing that $AS + SA^*$ is bounded and that G is invariant under $\mathcal{D}(A)$.

Lemma 3.4. *Let Assumptions 1 and 3 be satisfied. Then $AS + SA^* \in \mathcal{L}(H)$.*

Proof. Since $AS \in \mathcal{L}(H)$, so is $(AS)^*$. Furthermore, $S = S^*$ by Assumption 1, and we have $(AS)^*x = S^*A^*x = SA^*x$ for all $x \in \mathcal{D}(A^*)$. But $\mathcal{D}(A^*)$ is dense in H, e.g. by [15, Lemma 1.10.5], and hence $(AS)^* = SA^*$ on $\mathcal{L}(H)$. □

Remark 3.2. The conclusion of Lemma 3.4 is what is needed for the analysis, and could be directly assumed instead of the stronger condition given by Assumption 3. However, this would be more difficult to verify, and Assumption 3 more clearly shows that S is expected to be smoothing.

Lemma 3.5. *Let Assumptions 1 and 3 be satisfied and let $P \in \mathcal{D}(\mathcal{A})$. Then $GP \in \mathcal{D}(\mathcal{A})$.*

Proof. The operator GP is in the domain of $\mathcal{A}$ if $-GP$ is, and this is equivalent to the associated operator $\varphi_{PSP}(x, y)$ being continuous with respect to the norm on H. By Remark 2.1 it suffices to show that $\varphi_{PSP(x,y)}$ is continuous for $x, y \in \mathcal{D}(A)$. We have

$$\begin{aligned}
\varphi_{PSP}(x, y) &= (PSPx, Ay) + (Ax, PSPy) \\
&= (PSPx, Ay) + (Ax, PSPy) + (ASPx, Py) + (Px, ASPy) \\
&\quad - (ASPx, Py) + (Px, ASPy),
\end{aligned}$$

since AS is an element of $\mathcal{L}(H)$ by Assumption 3. Now if we let $x' = SPx$ and $y' = SPy$ we find

$$\begin{aligned}
\varphi_{PSP}(x, y) &= (Px', Ay) + (Ax, Py') + (Px, Ay') + (Ax', Py) \\
&\quad - ((AS + SA^*)Px, Py) \\
&= \varphi_P(x', y) + \varphi_P(x, y') - ((AS + SA^*)Px, Py).
\end{aligned}$$

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

$$\begin{aligned}
|\varphi_{PSP}(x, y)| &\leq |\varphi_P(x', y)| + |\varphi_P(x, y')| + \|AS + SA^*\|_{\mathcal{L}(H)}\|P\|^2_{\mathcal{L}(H)}\|x\|\|y\| \\
&\leq C(\|x'\|\|y\| + \|x\|\|y'\|) + C\|x\|\|y\| \\
&\leq 2C\|P\|_{\mathcal{L}(H)}\|S\|_{\mathcal{L}(H)}\|x\|\|y\| + C\|x\|\|y\|.
\end{aligned}$$

Thus φ_{PSP} is continuous with respect to the norm of H, so $GP \in \mathcal{D}(\mathcal{A})$. $\qquad \square$

Bounding the rest term $R(k\tau, x)$ is the key step in the proof of Lemma 3.2. The argument used there relies on $\dot{P}(t)x$ being uniformly bounded, which does not necessarily hold under Assumption 3 instead of 2. The next lemma shows that under Assumption 3 we can express the integrand of $R(k\tau, x)$ in terms of only $P(t)x$ and thereby avoid this issue.

Lemma 3.6. *Let Assumptions 1 and 3 be satisfied and let $P \in C_p\big([0, T], \Sigma(H)\big)$ be the solution to (2.1). Then for $r \in [0, \tau)$ and $t + r \in (0, T]$ the identification*

$$\frac{\mathrm{d}}{\mathrm{d}r}\big(\mathrm{e}^{(\tau - r)\mathcal{A}}GP(t + r)\big) = \mathrm{e}^{(\tau - r)\mathcal{A}}f(t + r),$$

holds, where $f : [0, T] \to \mathcal{L}(H)$ is defined by

$$f(s) = -QSP(s) - P(s)SQ + 2P(s)SP(s)SP(s) - P(s)(AS + SA^*)P(s). \quad (3.6)$$

As a consequence,

$$\big\|R(t, x)\big\|_{\mathcal{L}(H)} \leq C\tau^2\|x\| \qquad (3.7)$$

where $C = C(\gamma, \|AS\|_{\mathcal{L}(H)})$.

Proof. By Lemma 2.3, the solution P is an element in $C^\infty\big([\varepsilon,T],\Sigma(H)\big)$ and $P(t)\in\mathcal{D}(\mathcal{A})$ for all $t\in[\varepsilon,T]$, with $\varepsilon>0$. Thus, like in the proof of Lemma 3.1, for $t+r>0$ with $0\le r<\tau$ we have

$$\frac{\mathrm{d}}{\mathrm{d}r}\big(\mathrm{e}^{(\tau-r)\mathcal{A}}GP(t+r)\big)=\lambda_1(t,r)+\lambda_2(t,r),$$

where λ_j, $j=1,2$, are stated in Lemma 3.1. In the following, P and $\dot{P}$ are always evaluated at $t+r$, but we do not always write this out for readability reasons. For the same reason, we drop the arguments for the terms $\lambda_j(t,r)$.

By using the differential equation (2.1) to expand $\dot{P}(t+r)$ and writing one of the operator semigroups in terms of $\mathrm{e}^{\tau A}$ and $\mathrm{e}^{\tau A^*}$, we get

$$\lambda_2=-\mathrm{e}^{(\tau-r)A^*}\big((\mathcal{A}P)SP+PS(\mathcal{A}P)\big)\mathrm{e}^{(\tau-r)A}$$
$$-\mathrm{e}^{(\tau-r)\mathcal{A}}\big(QSP+PSQ-2PSPSP\big)=\lambda_{21}+\lambda_{22}.$$

Due to Lemma 3.5, we can expand the operator $\mathcal{A}$ in λ_{21} in terms of A and A^* and then reassemble it again with a different argument, as follows. First let $x\in\mathcal{D}(A)$. Then also $\mathrm{e}^{(\tau-r)A}x\in\mathcal{D}(A)$ and

$$\lambda_{21}x=-\mathrm{e}^{(\tau-r)A^*}\big((A^*P+PA)SP+PS(A^*P+PA)\big)\mathrm{e}^{(\tau-r)A}x$$
$$=-\mathrm{e}^{(\tau-r)A^*}(A^*PSP+PSPA+PASP+PSA^*P)\mathrm{e}^{(\tau-r)A}x$$
$$=-\mathrm{e}^{(\tau-r)A^*}\big(\mathcal{A}(PSP)+P(AS+SA^*)P\big)\mathrm{e}^{(\tau-r)A}x$$
$$=-\mathcal{A}\mathrm{e}^{(\tau-r)\mathcal{A}}PSPx-\mathrm{e}^{(\tau-r)\mathcal{A}}P(AS+SA^*)Px.$$

In the final equality, we have used the commutativity of $\mathcal{A}$ and $\mathrm{e}^{t\mathcal{A}}$. Since the operator $\mathcal{A}\mathrm{e}^{(\tau-r)\mathcal{A}}PSP-\mathrm{e}^{(\tau-r)\mathcal{A}}P(AS+SA^*)P$ is well defined in $\mathcal{L}(H)$ by Lemmas 3.5 and 3.4, and $\mathcal{D}(A)$ is dense in H (Remark 2.1), the above sequence of equalities hold for all $x\in H$. We thus have

$$\lambda_{21}=\mathcal{A}\mathrm{e}^{(\tau-r)\mathcal{A}}PSP-\mathrm{e}^{(\tau-r)\mathcal{A}}P(AS+SA^*)P$$

in the sense of $\mathcal{L}(H)$-operators. We now observe that the first term in λ_{21} is in fact λ_1, but with opposite sign. Thus,

$$\frac{\mathrm{d}}{\mathrm{d}r}\big(\mathrm{e}^{(\tau-r)\mathcal{A}}GP(t+r)\big)=\lambda_{22}-\mathrm{e}^{(\tau-r)\mathcal{A}}P(t+r)(AS+SA^*)P(t+r),$$

which is precisely $\mathrm{e}^{(\tau-r)\mathcal{A}}f(t+r)$. This proves the first claim.

Since $\|AS+SA^*\|_{\mathcal{L}(H)}\le C$ by Lemma 3.4, we then directly get by Lemma 2.2 that the integrand in (3.7) satisfies

$$\|\mathrm{e}^{(\tau-r)\mathcal{A}}f(t+r)\|_{\mathcal{L}(H)}\le C(\gamma,\|AS\|_{\mathcal{L}(H)}).$$

Since this bound is independent of r, and $\tau\le T$,

$$\|R(t,x)\|=\left\|\int_0^\tau\int_0^s\frac{\mathrm{d}}{\mathrm{d}r}\Big(\mathrm{e}^{(\tau-r)\mathcal{A}}GP(t+r)x\Big)\,\mathrm{d}r\mathrm{d}s\right\|$$
$$\le\int_0^\tau\int_0^s\|\mathrm{e}^{(\tau-r)\mathcal{A}}f(t+r)x\|\,\mathrm{d}r\mathrm{d}s$$
$$\le C(\gamma,\|AS\|_{\mathcal{L}(H)})\tau^2\|x\|,$$

which proves the second claim. $\qquad\square$

Lemma 3.7. *Let Assumptions 1 and 3 be fulfilled and let P be the solution to (2.1). If $\|\mathcal{L}_\tau^k P_0\|_{\mathcal{L}(H)} \leq \gamma + 1$ for all $k = 0, \ldots, n$, then*

$$\|P\big((n+1)\tau\big) - \mathcal{L}_\tau^{n+1} P_0\|_{\mathcal{L}(H)} \leq D\tau,$$

where $D = D(\gamma, \|AS\|_{\mathcal{L}(H)})$.

Proof. By following exactly the same approach as in the proof of Lemma 3.2, we find that
$$\|E_{n+1}x\| = K_1 + K_2 + K_3, \tag{3.8}$$
where K_j is defined in (3.5) and where

$$K_1 + K_2 \leq C(\gamma)\,\tau\big(\sum_{k=1}^{n}\|E_k\|_{\mathcal{L}(H)} + 1\big)\|x\|.$$

Furthermore, by Lemmas 2.2 and 3.6 and the trivial bound $k\tau \leq T$, we obtain

$$K_3 = \left\|\sum_{k=0}^{n} \mathrm{e}^{\tau(n-k)\mathcal{A}} R(k\tau, x)\right\| \leq C(\gamma, \|AS\|_{\mathcal{L}(H)})\,\tau\|x\|.$$

Summing up yields

$$\|E_{n+1}\|_{\mathcal{L}(H)} \leq \sup_{\|x\|=1} K_1 + K_2 + K_3 \leq C\tau\big(\sum_{k=0}^{n}\|E_k\|_{\mathcal{L}(H)} + 1\big),$$

where $C = C(\gamma, \|AS\|_{\mathcal{L}(H)})$, and the sought after error bound again follows by a discrete Grönwall inequality [6, Theorem 4.1]. $\qquad\square$

Lemma 3.7 together with the proof of Theorem 3.3 then give us our second convergence result:

Theorem 3.8. *Let Assumptions 1 and 3 be fulfilled and let P be the solution to (2.1). Then*

$$\|P(n\tau) - \mathcal{L}_\tau^n P_0\|_{\mathcal{L}(H)} \leq D\tau \quad \text{for all } \tau \leq \tau^*,$$

where the constants D, τ^* only depend on the solution $(\|P\|_\infty)$ and the problem data $(T, \|S\|_{\mathcal{L}(H)}, \|Q\|_{\mathcal{L}(H)}, \|AS\|_{\mathcal{L}(H)})$.

Remark 3.3. We note that unlike in the proof of Theorem 3.3 we did not use the bound on $\mathcal{A}\mathrm{e}^{t\mathcal{A}}$. This was possible because the terms can be rearranged such that all direct applications of A occur in $AS + SA^*$, which is bounded under Assumption 3.

4 Convergence analysis for Strang splitting

In this section, we analyze the Strang splitting scheme for (2.1) under the Assumptions 1 to 3. We consider an equidistant grid $t_n = n\tau$ with time step $\tau = T/N$ and approximate $P^n \approx P(t_n)$. In the notation of Section 2, the method is then defined by $P^{n+1} = \mathcal{S}_\tau P^n$, where the time-stepping operator $\mathcal{S}_\tau$ is given by

$$\mathcal{S}_\tau = \mathrm{e}^{\tau/2F}\mathrm{e}^{\tau G}\mathrm{e}^{\tau/2F}.$$

A second-order Taylor expansion of the nonlinear flow in $\mathcal{S}_\tau$ gives the representation

$$\begin{aligned}
\mathcal{S}_\tau P &= \mathrm{e}^{\tau\mathcal{A}}P + \int_0^\tau \mathrm{e}^{(\tau-s)\mathcal{A}}Q\,\mathrm{d}s + \tau\mathrm{e}^{\tau/2\mathcal{A}}(G + \tau M)\mathrm{e}^{\tau/2F}P \\
&\quad + \mathrm{e}^{\tau/2\mathcal{A}}\int_0^\tau \frac{(\tau-s)^2}{2}\frac{\mathrm{d}^3}{\mathrm{d}s^3}\mathrm{e}^{sG}\mathrm{e}^{\tau/2F}P\,\mathrm{d}s \\
&=: S_0(P) + S_1 + S_2(P) + S_3(P)
\end{aligned} \tag{4.1}$$

in $\mathcal{L}(H)$, where M is defined in Lemma 2.8. Next, we reformulate the exact solution. As for the reformulation in the Lie case, we make use of the fact that the mapping $s \mapsto \mathrm{e}^{(\tau-s)\mathcal{A}}GP(s)$ is in $C^\infty\big([\varepsilon, \tau-\varepsilon], \Sigma(H)\big)$ and obtain the reformulation

$$\begin{aligned}
P(\tau) &= P(\tau - \varepsilon + \varepsilon) \\
&= \mathrm{e}^{(\tau-\varepsilon)\mathcal{A}}P(\varepsilon) + \int_\varepsilon^\tau \mathrm{e}^{(\tau-s)\mathcal{A}}\big(Q + GP(s)\big)\,\mathrm{d}s \\
&= \mathrm{e}^{(\tau-\varepsilon)\mathcal{A}}P(\varepsilon) + \int_\varepsilon^\tau \mathrm{e}^{(\tau-s)\mathcal{A}}Q\,\mathrm{d}s + \int_{\tau-\varepsilon}^\tau \mathrm{e}^{(\tau-s)\mathcal{A}}GP(s)\,\mathrm{d}s \\
&\quad + \mathrm{e}^{\tau/2\mathcal{A}}GP(\tau/2)\int_\varepsilon^{\tau-\varepsilon} 1\,\mathrm{d}s + \frac{\mathrm{d}}{\mathrm{d}r}\Big(\mathrm{e}^{(\tau-r)\mathcal{A}}GP(r)\Big)\Big|_{r=\tau/2}\int_\varepsilon^{\tau-\varepsilon}(s - \tau/2)\,\mathrm{d}s \\
&\quad + \int_\varepsilon^{\tau-\varepsilon}\int_{\tau/2}^s (s-r)\frac{\mathrm{d}^2}{\mathrm{d}r^2}\Big(\mathrm{e}^{(\tau-r)\mathcal{A}}GP(r)\Big)\,\mathrm{d}r\,\mathrm{d}s \\
&= I_0(\varepsilon, P_0) + I_1(\varepsilon) + I_2(\varepsilon, P_0) + I_3(\varepsilon, P_0) + I_4(\varepsilon, P_0) + I_5(\varepsilon, P_0).
\end{aligned}$$

From this we find that for any $x \in H$ the first two terms match the expansion of the method, i.e.

$$\lim_{\varepsilon\to 0} I_0(\varepsilon, P_0)x = S_0(P_0)x \quad\text{and}\quad \lim_{\varepsilon\to 0} I_1(\varepsilon)x = S_1 x.$$

The integrand of I_2 is uniformly bounded on $[\tau - \varepsilon, \tau]$ and the integrand of I_4 is symmetric, which imply that

$$\lim_{\varepsilon\to 0} I_j(\varepsilon, P_0)x = 0 \quad\text{for } j = 2, 4.$$

Finally, we note that

$$I_3(P_0) = \lim_{\varepsilon\to 0} I_3(\varepsilon, P_0) = \tau\mathrm{e}^{\tau/2\mathcal{A}}GP(\tau/2).$$

Thus we conclude that $\lim_{\varepsilon \to 0} I_5(\varepsilon, P_0)x$ exists and equals $R(0, x)$ where

$$R(t, x) = \int_0^\tau \int_{\tau/2}^s (s - r) \frac{\mathrm{d}2}{\mathrm{d}r^2} \Big(\mathrm{e}^{(\tau-s)\mathcal{A}} GP(t + r) \Big) x \, \mathrm{d}r \mathrm{d}s.$$

This reformulation of the exact solution can be made not only at $t = 0$, but around any $t = k\tau \leq T - \tau$ resulting in

$$P((k + 1)\tau)x = S_0(P(k\tau))x + S_1 x + I_3(P(k\tau))x + R(k\tau, x).$$

Throughout this section we once more choose a $\gamma > 0$ such that

$$\max\{\|P\|_\infty, \|\dot{P}\|_\infty\} \leq \gamma.$$

Before we prove consistency for the Strang splitting, we derive two preliminary results:

Lemma 4.1. *Let Assumptions 1 and 3 be fulfilled, then*

$$\|\mathrm{e}^{\tau A} S \mathrm{e}^{\tau A^*} - S\|_{\mathcal{L}(H)} \leq C\tau,$$

where $C = C(\|AS\|_{\mathcal{L}(H)})$.

Proof. By expressing the difference as an integral,

$$
\begin{aligned}
\|(\mathrm{e}^{\tau A} S \mathrm{e}^{\tau A^*} - S)x\| &= \left\| \int_0^\tau A \mathrm{e}^{sA} S \mathrm{e}^{sA^*} x + \mathrm{e}^{sA} S A^* \mathrm{e}^{sA^*} x \, \mathrm{d}s \right\| \\
&= \left\| \int_0^\tau \mathrm{e}^{sA} A S \mathrm{e}^{sA^*} x + \mathrm{e}^{sA} S A^* \mathrm{e}^{sA^*} x \, \mathrm{d}s \right\| \\
&\leq C\big(\|AS\|_{\mathcal{L}(H)} + \|SA^*\|_{\mathcal{L}(H)} \big) \tau \|x\|.
\end{aligned}
$$

Note that commuting A and e^{sA} is justified, since AS is a bounded operator. $\square$

Lemma 4.2. *Let Assumptions 1 to 3 be satisfied and let P be the solution to (2.1). Then, with f as in Lemma 3.6, the equality*

$$\frac{\mathrm{d}2}{\mathrm{d}r^2} \big(\mathrm{e}^{(\tau-r)\mathcal{A}} GP(t + r) \big) = -\mathcal{A}\mathrm{e}^{(\tau-r)\mathcal{A}} f(t + r) + \mathrm{e}^{(\tau-r)\mathcal{A}} \frac{\mathrm{d}}{\mathrm{d}r} f(t + r),$$

is valid for $r \in [0, \tau)$ and $t + r \in (0, T]$. Furthermore,

$$\left\| \int_0^\tau \int_{\tau/2}^s (s - r) \mathrm{e}^{(\tau-r)\mathcal{A}} \frac{\mathrm{d}}{\mathrm{d}r} f(t + r) x \, \mathrm{d}r \mathrm{d}s \right\| \leq C\tau^3 \|x\|$$

with a constant $C = C(\gamma, \|AS\|_{\mathcal{L}(H)})$.

Proof. As $0 \leq r < \tau$, $0 < t + r \leq T$ and Assumption 3 is fulfilled, Lemma 3.6 gives

$$\frac{\mathrm{d}}{\mathrm{d}r} \Big(\mathrm{e}^{(\tau-r)\mathcal{A}} GP(t + r) \Big) = \mathrm{e}^{(\tau-r)\mathcal{A}} f(t + r).$$

Since f is an element of $C^\infty\big([\varepsilon, T], \Sigma(H)\big)$, as it is a product of elements in $\mathcal{L}(H)$ and the solution $P \in C^\infty\big([\varepsilon, T], \Sigma(H)\big)$, we obtain

$$\frac{\mathrm{d}}{\mathrm{d}r}\mathrm{e}^{(\tau-r)\mathcal{A}}f(r) = -\mathcal{A}\mathrm{e}^{(\tau-r)\mathcal{A}}f(r) + \mathrm{e}^{(\tau-r)\mathcal{A}}\frac{\mathrm{d}}{\mathrm{d}r}f(r),$$

i.e. the desired equality. Writing out the last term yields the identity

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}r}f(t+r) = & -QS\dot{P} - \dot{P}SQ + 2(\dot{P}SPSP + PS\dot{P}SP + PSPS\dot{P}) \\
& -\dot{P}(AS + SA^*)P - P(AS + SA^*)\dot{P},
\end{aligned}$$

and the bound

$$\sup_{r\in(0,\tau)}\left\|\mathrm{e}^{(\tau-r)\mathcal{A}}\frac{\mathrm{d}}{\mathrm{d}r}f(t+r)\right\|_{\mathcal{L}(H)} \leq C\left\|\frac{\mathrm{d}}{\mathrm{d}r}f\right\|_\infty \leq C(\gamma, \|AS\|_{\mathcal{L}(H)}). \qquad (4.2)$$

Hence,

$$\begin{aligned}
\Bigg\|\int_0^\tau & \int_{\tau/2}^s (s-r)\mathrm{e}^{(\tau-r)\mathcal{A}}\frac{\mathrm{d}}{\mathrm{d}r}f(r)x\,\mathrm{d}r\mathrm{d}s\Bigg\| \\
& \leq \int_0^{\tau/2}\int_s^{\tau/2}(s-r)\left\|\mathrm{e}^{(\tau-r)\mathcal{A}}\frac{\mathrm{d}}{\mathrm{d}r}f(t+r)x\right\|\mathrm{d}r\mathrm{d}s \\
& \quad + \int_{\tau/2}^\tau\int_{\tau/2}^s(s-r)\left\|\mathrm{e}^{(\tau-r)\mathcal{A}}\frac{\mathrm{d}}{\mathrm{d}r}f(t+r)x\right\|\mathrm{d}r\mathrm{d}s \\
& \leq C\left\|\frac{\mathrm{d}}{\mathrm{d}r}f\right\|_\infty\int_0^{\tau/2}\int_s^{\tau/2}(s-r)\,\mathrm{d}r\mathrm{d}s\|x\| \\
& \leq C(\gamma, \|AS\|_{\mathcal{L}(H)})\tau^3\|x\|,
\end{aligned}$$

and the desired inequality is shown. We note that the integral was split in two since the direction of integration in the inner integral changes when $s < \tau/2$. $\qquad\square$

Remark 4.1. Note that the expression for the second derivative consists of two terms, and we only bound the second term. The first term will be handled separately in the proof of Lemma 4.3.

We can now prove consistency for the Strang splitting:

Lemma 4.3. *Let Assumptions 1 to 3 be fulfilled and let P be the solution to* (2.1)*. If $\|\mathcal{S}_\tau^k P_0\|_{\mathcal{L}(H)} \leq \gamma + 1$ for all $k = 0, \ldots n$, then*

$$\|E_{n+1}\|_{\mathcal{L}(H)} \leq D\tau^2(1 + |\log\tau|),$$

where $D = D(\gamma, \|AS\|_{\mathcal{L}(H)})$.

Proof. Comparing the approximation and exact solution we get

$$\begin{aligned}
E_{n+1}x = {} & P((n+1)\tau)x - \mathcal{S}_\tau^{n+1}P_0x \\
= {} & \mathrm{e}^{\tau\mathcal{A}}\big(P(n\tau) - \mathcal{S}_\tau^n P_0\big)x \\
& + \big(S_2(P(n\tau)) - S_2(\mathcal{S}_\tau^n P_0)\big)x \\
& + \big(I_3 - S_2\big)(P(n\tau))x \\
& + R(n\tau, x) - S_3(\mathcal{S}_\tau^n(P_0))x.
\end{aligned}$$

21

Again we see that the first term is $e^{\tau \mathcal{A}} E_n x$ so this term can repeatedly be rewritten using the above equality, resulting in the bound

$$\|E_{n+1} x\| \le C \sum_{k=1}^{n} \|\big(S_2(P(k\tau)) - S_2(\mathcal{S}_\tau^k P_0)\big) x\|$$

$$+ C \sum_{k=0}^{n} \|(I_3 - S_2)(P(k\tau)) x\|$$

$$+ \left\| \sum_{k=0}^{n} e^{(n-k)\tau \mathcal{A}} R(k\tau, x) \right\| + C \sum_{k=0}^{n} \|S_3(\mathcal{S}_\tau^k P_0) x\|$$

$$=: K_1 + K_2 + K_3 + K_4.$$

Applying Lemmas 2.2 and 2.8 yields

$$K_1 \le C(\gamma)\tau(1+\tau) \sum_{k=1}^{n} \|E_n\|_{\mathcal{L}(H)} \|x\|.$$

The term K_4 is similarly straightforward to bound, since the hypothesis implies that
$$\|e^{\tau/2 F} \mathcal{S}_\tau^k P_0 x\|_{\mathcal{L}(H)} \le C(\gamma) \quad \text{for } k = 0, \dots, n,$$

and Lemma 2.7 then gives the bound

$$K_4 \le C \sum_{k=0}^{n} \left\| s \mapsto \frac{d^3}{ds^3} e^{sG} e^{\tau/2 F} \mathcal{S}_\tau^k P_0 \right\|_\infty \int_0^\tau \frac{(\tau - s)^2}{2} \, ds \|x\| \le C(\gamma)\tau^2 \|x\|.$$

Next, we bound the K_2 term, starting by only considering a single summand. With the shorthand
$$\hat{P} = e^{\tau/2 F} P(k\tau)$$

we obtain

$$\|(I_3 - S_2)(P(k\tau)) x\| \le C\tau \|\big(GP((k + 1/2)\tau) - G\hat{P}\big) x + \tau M \hat{P} x\|. \qquad (4.3)$$

Introduce the function g defined by

$$g(t, s) = \int_0^s e^{(\tau/2 - r)\mathcal{A}} GP(t + r) \, dr,$$

where $r \mapsto g(t, r) \in C^\infty\big([\varepsilon, \tau/2 - \varepsilon], \Sigma(H)\big)$ for any $\varepsilon > 0$, with its first two derivatives being uniformly bounded by Lemma 3.6. This gives the representation $P((k + 1/2)\tau) = \hat{P} + g(k\tau, \tau/2)$ and the equality

$$GP((k + 1/2)\tau) x - G\hat{P} x = \int_0^{\tau/2} \frac{d}{ds} G(\hat{P} + g(k\tau, s)) x \, ds$$

$$= -\frac{\tau}{2} \big(g_s(k\tau, 0) S\hat{P} + \hat{P} S g_s(k\tau, 0)\big) x + \int_0^{\tau/2} \int_0^s \frac{d^2}{dr^2} G(\hat{P} + g(k\tau, r)) x \, dr ds,$$

where $g_s = \partial g / \partial s$. Inserting this equality into (4.3) and summing up yields

$$K_2 \le C \sum_{k=0}^{n} \tau^2 \left\| \big(g_s(k\tau,0)S\hat{P} + \hat{P}Sg_s(k\tau,0) - 2M\hat{P}\big)x \right\|$$
$$+ C \sum_{k=0}^{n} \tau \left\| \int_0^{\tau/2} \int_0^s \frac{\mathrm{d}^2}{\mathrm{d}r^2} G\big(\hat{P} + g(k\tau,r)\big)x \,\mathrm{d}r\mathrm{d}s \right\| =: K_{21} + K_{22}.$$

Now let P be shorthand for $P(k\tau)$. For the individual K_{21}-summands, writing out $\hat{P} = \mathrm{e}^{\tau/2\mathcal{A}}P + \int_0^{\tau/2} \mathrm{e}^{(\tau/2-s)\mathcal{A}}Q\,\mathrm{d}s$ results in many terms, most of which contain the integral as a factor. Since all the other factors in these terms are uniformly bounded, they are all bounded by $C(\gamma)\tau$. We thus get

$$\left\| \big(g_s(k\tau,0)S\hat{P} + \hat{P}Sg_s(k\tau,0) - 2M\hat{P}\big)x \right\|$$
$$= \| \big(\mathrm{e}^{\tau/2\mathcal{A}}(PSP)S\hat{P}x + \hat{P}S\mathrm{e}^{\tau/2\mathcal{A}}(PSP) - 2\hat{P}S\hat{P}S\hat{P}\big)x \|$$
$$\le \| \big(G\big(\mathrm{e}^{\tau/2\mathcal{A}}P\big)S\mathrm{e}^{\tau/2\mathcal{A}}P - \mathrm{e}^{\tau/2\mathcal{A}}(GP)S\mathrm{e}^{\tau/2\mathcal{A}}P\big)x \|$$
$$\quad + \| \Big(\big(\mathrm{e}^{\tau/2\mathcal{A}}P\big)SG\big(\mathrm{e}^{\tau/2\mathcal{A}}P\big) - \big(\mathrm{e}^{\tau/2\mathcal{A}}P\big)S\mathrm{e}^{\tau/2\mathcal{A}}(GP)\Big)x \| + C(\gamma)\tau$$
$$\le \tau C(\gamma, \|AS\|_{\mathcal{L}(H)})\|x\|,$$

where the last bound follows from Lemma 4.1.

Furthermore, the integrand of each K_{22}-summand can be bounded by

$$\left\| \frac{\mathrm{d}^2}{\mathrm{d}r^2} G\big(\hat{P} + g(k\tau,r)\big) \right\|_{\mathcal{L}(H)}$$
$$= \left\| \big(\hat{P} + g(k\tau,r)\big)S\mathrm{e}^{(\tau/2-r)\mathcal{A}}f(k\tau+r) + \mathrm{e}^{(\tau/2-r)\mathcal{A}}f(k\tau+r)S\big(\hat{P} + g(k\tau,r)\big) \right.$$
$$\left. + 2\mathrm{e}^{(\tau/2-r)\mathcal{A}}\big(PSP\big)S\mathrm{e}^{(\tau/2-r)\mathcal{A}}\big(PSP\big) \right\|_{\mathcal{L}(H)}$$
$$\le 2C\|S\|_{\mathcal{L}(H)}\Big(\|f\|_\infty \sup_{r\in(0,\tau/2)} \|g(k\tau,r)\| + \|f\|_\infty\|\hat{P}\|_{\mathcal{L}(H)} + \|P\|_\infty^4\|S\|_{\mathcal{L}(H)}^2\Big),$$

since $\frac{\partial^2 g(t,r)}{\partial r^2} = \mathrm{e}^{(\tau/2-r)\mathcal{A}}f(t+r)$. Applying the derived bounds on the summands of K_{2j} gives

$$K_2 \le C(\gamma, \|AS\|_{\mathcal{L}(H)})\tau^2\|x\|.$$

Finally we bound K_3. By Lemma 4.2 we obtain that

$$K_3 \le K_{31} + K_{32}$$

where

$$K_{3j} = \left\| \sum_{k=0}^{n} \mathrm{e}^{(n-k)\tau\mathcal{A}} \int_0^\tau \int_{\tau/2}^s \lambda_j(k\tau,r)x\,\mathrm{d}r\mathrm{d}s \right\| \quad \text{for} \quad j = 1,2,$$

with

$$\lambda_1(t,r) = -(s-r)\mathcal{A}\mathrm{e}^{(\tau-r)\mathcal{A}}f(t+r)$$
$$\lambda_2(t,r) = (s-r)\mathrm{e}^{(\tau-r)\mathcal{A}}\frac{\mathrm{d}}{\mathrm{d}r}f(t+r).$$

Note that the improper integral $\int_0^\tau \int_{\tau/2}^s \lambda_1(k\tau, r)x \, drds$ is well defined since $r \mapsto \lambda_2(t, r)$ is uniformly bounded by (4.2) and

$$\lim_{\varepsilon \to 0^+} \int_\varepsilon^{\tau - \varepsilon} \int_{\tau/2}^s \lambda_1(k\tau, r)x \, drds = R(t, x) - \int_0^\tau \int_{\tau/2}^s \lambda_2(k\tau, r)x \, drds.$$

From the bound in Lemma 4.2 we directly have

$$K_{32} \le C \sum_{k=0}^n \left\| \int_0^\tau \int_{\tau/2}^s \lambda_2(k\tau, r)x \, drds \right\| \le C(\gamma, \|AS\|_{\mathcal{L}(H)})\tau^2.$$

For K_{31} we proceed as in Section 3.1, bounding the $k = n$ term separately:

$$\left\| \int_0^\tau \int_{\tau/2}^s \lambda_1(n\tau, r)x \, drds \right\| \le \int_0^{\tau/2} \int_s^{\tau/2} \|\lambda_1(n\tau, r)x\| \, drds$$

$$+ \int_{\tau/2}^\tau \int_{\tau/2}^s \|\lambda_1(n\tau, r)x\| \, drds$$

$$\le C\|f\|_\infty \left(\int_0^{\tau/2} \int_s^{\tau/2} \frac{r - s}{\tau - r} \, drds + \int_{\tau/2}^\tau \int_{\tau/2}^s \frac{s - r}{\tau - r} \, drds \right) \|x\|$$

$$\le C(\gamma, \|AS\|_{\mathcal{L}(H)}) \int_{\tau/2}^\tau \int_{\tau/2}^s \frac{s - r}{\tau - r} \, drds \|x\| = C(\gamma, \|AS\|_{\mathcal{L}(H)})\tau^2 \|x\|.$$

For the rest of the sum we have the bound

$$\left\| \sum_{k=0}^{n-1} e^{(n-k)\tau \mathcal{A}} \int_0^\tau \int_{\tau/2}^s \lambda_1(k\tau, r)x \, drds \right\|$$

$$\le \sum_{k=0}^{n-1} \int_0^{\tau/2} \int_s^{\tau/2} \|(\tau/2 - r)e^{(\tau-r)\mathcal{A}} \mathcal{A}e^{(n-k)\tau \mathcal{A}} f(t + r)x\| \, drds$$

$$+ \sum_{k=0}^{n-1} \int_{\tau/2}^\tau \int_{\tau/2}^s \|(\tau/2 - r)e^{(\tau-r)\mathcal{A}} \mathcal{A}e^{(n-k)\tau \mathcal{A}} f(t + r)x\| \, drds$$

$$\le C\|f\|_\infty \left(\sum_{k=0}^{n-1} \int_{\tau/2}^\tau \int_{\tau/2}^s \frac{s - r}{(n - k)\tau} \, drds \right) \|x\|$$

$$\le C(\gamma, \|AS\|_{\mathcal{L}(H)})\tau^3 \sum_{k=0}^{n-1} \frac{1}{(n - k)\tau} \|x\|$$

$$\le C(\gamma, \|AS\|_{\mathcal{L}(H)})\tau^2 (1 + \log n) \|x\|.$$

Combining the bounds for K_j give

$$\|E_{n+1}\|_{\mathcal{L}(H)} = \sup_{\|x\|=1} K_1 + K_2 + K_3 + K_4$$

$$\le C(\gamma)\tau \sum_{k=1}^n \|E_k\|_{\mathcal{L}(H)} + C(\gamma, \|AS\|_{\mathcal{L}(H)})\tau^2 (1 + |\log \tau|).$$

From this the desired bound is found by a discrete Grönwall inequality. $\qquad \square$

The convergence analysis of the Strang splitting scheme now follows by an identical argument to the one in the proof of Theorem 3.3, using the bound provided by Lemma 4.3 instead of Lemma 3.2:

Theorem 4.4. *Let Assumptions 1 to 3 be fulfilled and let P be the solution to (2.1). Then*

$$\|P(n\tau) - \mathcal{S}_\tau^n(P_0)\|_{\mathcal{L}(H)} \leq D\tau^2(1 + |\log \tau|) \quad \text{for all } \tau \leq \tau^*,$$

where the constants D, τ^* depend only on the solution ($\|P\|_\infty, \|\dot{P}\|_\infty$) and problem data $(T, \|S\|_{\mathcal{L}(H)}, \|Q\|_{\mathcal{L}(H)}, \|AS\|_{\mathcal{L}(H)})$.

5 Numerical experiments

In order to illustrate the theoretical results, we now present several numerical experiments. They are all based on the connection between DREs and the optimal control of LQR problems mentioned in the introduction. We consider the problem of minimizing the cost functional

$$J(u) = \int_0^T \|y(t)\|^2 + \|u(t)\|^2 \, \mathrm{d}t + \|Gx(T)\|^2, \tag{5.1}$$

subject to

$$\begin{cases} \dot{x} = Ax + Bu & \text{in } \Omega \\ y = Ex, \end{cases} \tag{5.2}$$

where $\Omega = (0,1)$. We set $H = L^2(\Omega)$ and choose $A = \Delta$, the Laplacian with periodic boundary conditions, and $Ex = \int_\Omega x$, such that the problem is a controlled heat equation on the unit interval where we observe the mean of the state x. The control operator $B : \mathbb{R} \to H$ and the initial condition $P_0 = G^*G : H \to H$ will be chosen such that different combinations of Assumptions 2 and 3 are fulfilled.

A family of examples that fulfill Assumption 3 can easily be constructed by letting $B : \mathbb{R} \to H$ be given by

$$Bu = u\rho$$

for any $\rho \in W \subset H$. The adjoint $B^* : H \to \mathbb{R}$ is given by

$$B^*v = (v, \rho)_H,$$

so $S = BB^* : H \to H$ satisfies

$$Sv = BB^*v = (\rho, v)_H \rho.$$

This means that S actually maps into W. By choosing $W = \mathcal{D}(A) = H^2 \cap H^1_{\mathrm{per}}(\Omega)$, Assumption 3 is satisfied. Using the same construction but instead choosing $\rho \in H \setminus \mathcal{D}(A)$ means that the assumption will not be satisfied. Since

Experiment	S	P_0	Assumption 2	Assumption 3
1	ρ_1	ρ_2	Yes	Yes
2	ρ_1	ξ_2	No	Yes
3	ξ_1	ρ_2	Yes	No
4	ξ_1	ξ_2	No	No

Table 1: Overview of the experiment setup.

$A = A^*$ in our experiments, the above procedure can also be used to construct a P_0 that either satisfies Assumption 2 or not, by simply replacing B with G^*.

We generate the functions ρ_1 and ρ_2 in $H^2 \cap H^1_{\mathrm{per}}(\Omega)$ and ξ_1 and ξ_2 in $H^0(\Omega) = H$ by sampling two different Q-Wiener processes, as described in [11, Section 10.2]. Essentially, this means that we compute weighted sums of eigenfunctions to the Laplacian, where the weights are appropriately scaled random numbers. The functions are initially evaluated on a grid with $N_h = 2^{17}$ grid points and then projected onto each of the coarser grids. We choose different combinations of ρ_j and ξ_j, $j = 1, 2$, to construct S and P_0 as described above, such that all combinations of fulfilling Assumption 2 and Assumption 3 are covered. These choices are listed in Table 1, and the functions themselves are plotted in Figure 1.

Remark 5.1. The operators $S = 0$ and $P_0 = 0$ clearly satisfy the assumptions. However, $S = 0$ is not a good choice for our purposes here, since it completely removes the nonlinearity in the DRE and thus means that the splitting schemes are exact. Similarly, $P_0 = 0$ leads to an exact solution P which is $C^\infty\big([0, T], \Sigma(H)\big)$. This follows from a minor modification to the proof of Theorem IV-1.3.1 in [5]. While it has not been rigorously proven in this paper, we expect that such high regularity will lead to optimal convergence orders for both schemes, regardless of any assumption on S. That is, Assumption 3 would not need to be satisfied for the Strang splitting to be of order 2 when $P_0 = 0$. This is why we do not use this initial condition in the experiments and instead consider a more general one.

Since we cannot time-step the operator-valued DREs directly, we first discretize in space using the finite element method (FEM) with piecewise linear basis functions φ_j, $j = 1, \ldots, N_h$. We denote the approximation space by $V_h := \mathrm{span}\big(\{\varphi_j\}_{j=1}^{N_h}\big)$ and note that $V_h \subset V = H^1_{\mathrm{per}}(\Omega) \subset H$. As described in detail in [12], the operator-valued DRE (2.1) then turns into the matrix-valued DRE

$$\boldsymbol{M\dot{P}M} = \boldsymbol{MPA} + \boldsymbol{A}^T\boldsymbol{PM} + \boldsymbol{E}^T\boldsymbol{E} - \boldsymbol{MPBB}^T\boldsymbol{PM}, \qquad (5.3)$$

where the matrices $\boldsymbol{A} \in \mathbb{R}^{N_h \times N_h}$, $\boldsymbol{M} \in \mathbb{R}^{N_h \times N_h}$, $\boldsymbol{B} \in \mathbb{R}^{N_h \times 1}$, $\boldsymbol{E} \in \mathbb{R}^{1 \times N_h}$ are given by $\boldsymbol{A}_{i,j} = (A\varphi_j, \varphi_i)$, $\boldsymbol{M}_{i,j} = (\varphi_j, \varphi_i)$, $\boldsymbol{B}_{i,j} = (B1, \varphi_i)$, $\boldsymbol{E}_{i,j} = (E\varphi_j, 1)$. The matrix $\boldsymbol{P}(t) \in \mathbb{R}^{N_h \times N_h}$ is the matrix representation of the FEM-approximation $P_h(t) : V_h \to V_h$ to $P(t) : H \to H$. If $V_h \ni x = \sum_{j=1}^{N_h} \boldsymbol{x}_j \varphi_j^h$ with
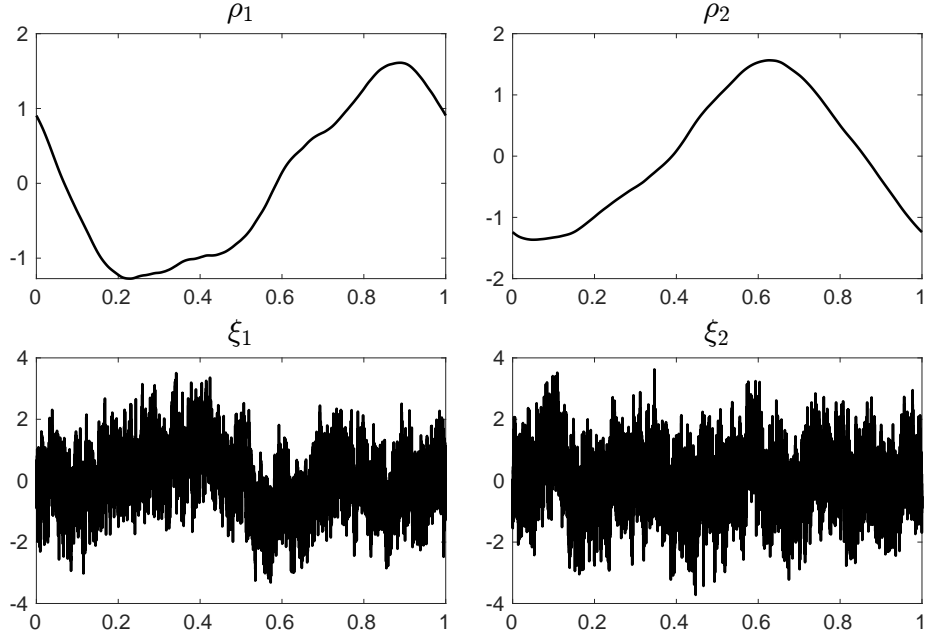
Figure 1: The functions $\rho_j \in H^2 \cap H^1_{\mathrm{per}}(\Omega)$ and $\xi_j \in H$, $j = 1, 2$ used to generate the operators P_0 and S in the experiments.

$\boldsymbol{x} \in \mathbb{R}^{N_h}$ then

$$P_h(t)x = \sum_{i,j=1}^{N} \boldsymbol{P}_{i,j}(t)(x, \varphi_j)\varphi_i. \tag{5.4}$$

The initial value $P_h(0)$ is the projection of $P(0)$ onto $\mathcal{L}(V_h, V_h)$, which satisfies $P_h(0) = (\mathrm{Id}_h)^* P(0) \, \mathrm{Id}_h$ where $\mathrm{Id}_h : V_h \to V \subset H$ is the identity operator and $(\mathrm{Id}_h)^* : H \to V_h$ is the H-orthogonal projection onto V_h.

The natural extension of $P_h(t)$ to an operator on H is given by $\mathrm{Id}_h \, P_h(t)(\mathrm{Id}_h)^*$, and the $\mathcal{L}(H)$-norm of this operator is given by

$$\|\mathrm{Id}_h \, P_h(t)(\mathrm{Id}_h)^*\|_{\mathcal{L}(H)} = \|\boldsymbol{L}_{\boldsymbol{M}}^T \boldsymbol{P}(t)\boldsymbol{L}_{\boldsymbol{M}}\|_{\mathbb{R}^{N \times N}},$$

where $\boldsymbol{M} = \boldsymbol{L}_{\boldsymbol{M}} \boldsymbol{L}_{\boldsymbol{M}}^T$ is a Cholesky factorization of the mass matrix and $\|\cdot\|_{\mathbb{R}^{N \times N}}$ denotes the standard spectral matrix norm.

In each experiment, we use several different spatial discretizations, corresponding to $N_h = 2^k$, $k = 2, \ldots, 14$. For each N_h, we apply the Lie and Strang splitting methods with $N_t = 2^j$ time steps for $j = 2, \ldots, 12$, corresponding to time steps of size $\tau = 2^{-j}$, until reaching $T = 0.1$. We use the low-rank-factored implementations of these schemes provided in the Matlab package DRESplit[1] with the low-rank tolerance set to 10^{-15}.

---

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

Denote the approximation of $P_h(n\tau)$ by P_h^n. The temporal errors for each h are computed as

$$\mathrm{err}_\tau = \|\mathrm{Id}_h(\dot{P}_h - \dot{P}_{\tau,\mathrm{ref}})(\mathrm{Id}_h)^*\|_\infty$$

where the reference approximation $P_{\tau,\mathrm{ref}}$ is given by applying the Strang splitting approximation with time step $\tau_{\mathrm{ref}} = 2^{-13}$ and discarding the intermediate steps which do not coincide with $n\tau$, $n = 1, \ldots, N_t$.

As N_h increases, the spatial discretization gets finer, and the matrix-valued DRE tends to the corresponding operator-valued DRE as $N_h \to \infty$. For each fixed spatial discretization, we expect the time discretizations to converge with order 1 and 2, respectively. But when $N_h \to \infty$, this should break down, unless a relevant combination of Assumptions 2 and 3 are fulfilled. One way this can manifest in is that the largest τ for which the "correct" orders are observed decreases as $N_h \to \infty$.

5.1 Experiment 1

The errors are shown in Figure 2, and we observe temporal convergence with perfect order 1 and 2 for the Lie and Strang splittings, respectively, regardless of the spatial discretization. Since both Assumptions 2 and 3 are fulfilled, this is the expected result according to Theorems 3.3, 3.8 and 4.4.

5.2 Experiment 2

The errors are shown in Figure 3, and we again observe temporal convergence with essentially perfect order 1 for the Lie splitting, regardless of the spatial discretization. This is in line with Theorem 3.8. For the Strang splitting, however, only the coarsest spatial discretization, corresponding to $N_h = 4$, yields a curve of slope 2 for the whole interval of time steps. For finer spatial discretizations, the errors behave more like $\mathcal{O}(\tau)$ until τ is small enough. Since only one of the assumptions required for Theorem 4.4 is satisfied, this is the expected result. We also note that, in contrast to Experiment 1, the errors initially increase as N_h increases. It is unclear why this increase stops around $N_h = 128$, but we conjecture that is is related to the fact that each spatial discretization employs its own reference approximation.

5.3 Experiment 3

The errors in this experiment are shown in Figure 4, and they are quite similar to those in Experiment 1. Since Assumption 2 is satisfied, Theorem 3.3 explains the Lie results. However, as Assumption 3 is not fulfilled, Theorem 4.4 cannot be used to argue for why the Strang splitting performs so well. With this said, we have not proven that the given assumptions are necessary conditions, only sufficient. Better than expected convergence properties can occur in some cases, but this cannot be relied upon to always happen.
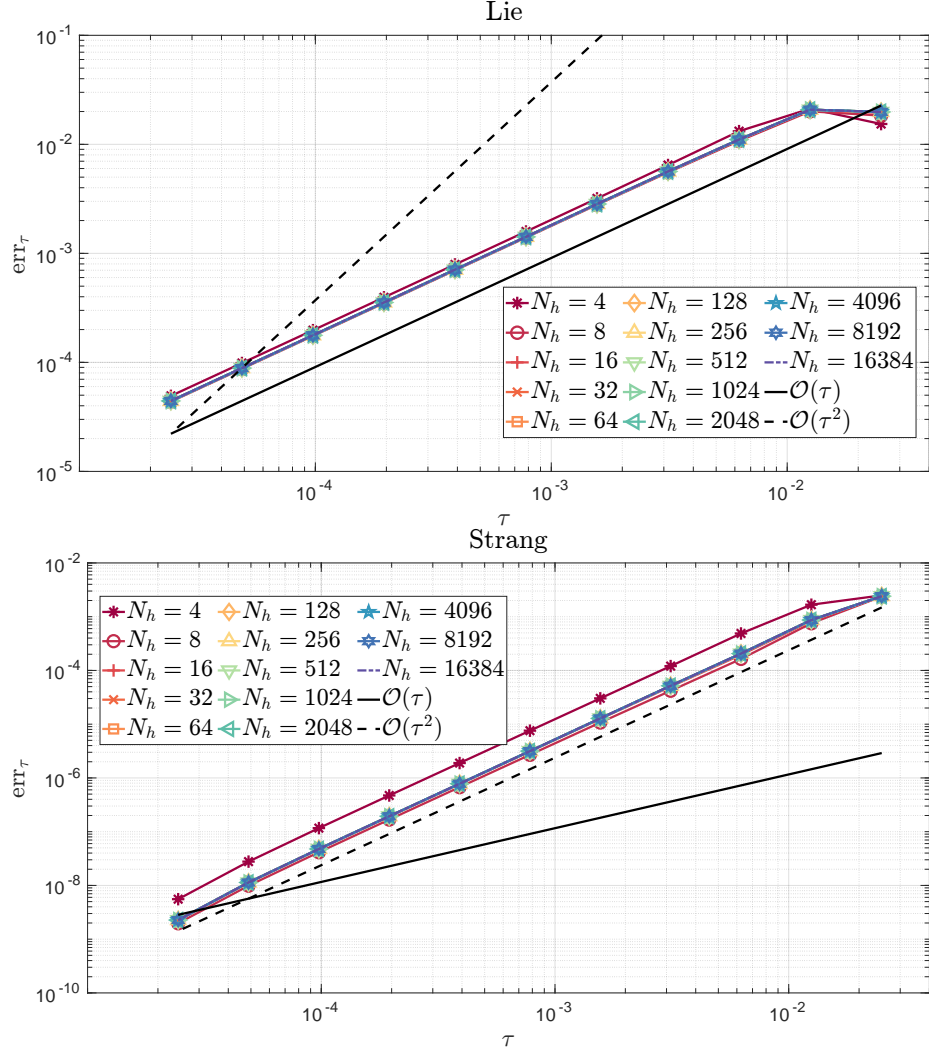
Figure 2: The errors err_τ for Lie (top) and Strang (bottom) splitting in a setting when Assumptions 2 and 3 are both fulfilled. Each curve corresponds to a different spatial discretization with N_h nodes.
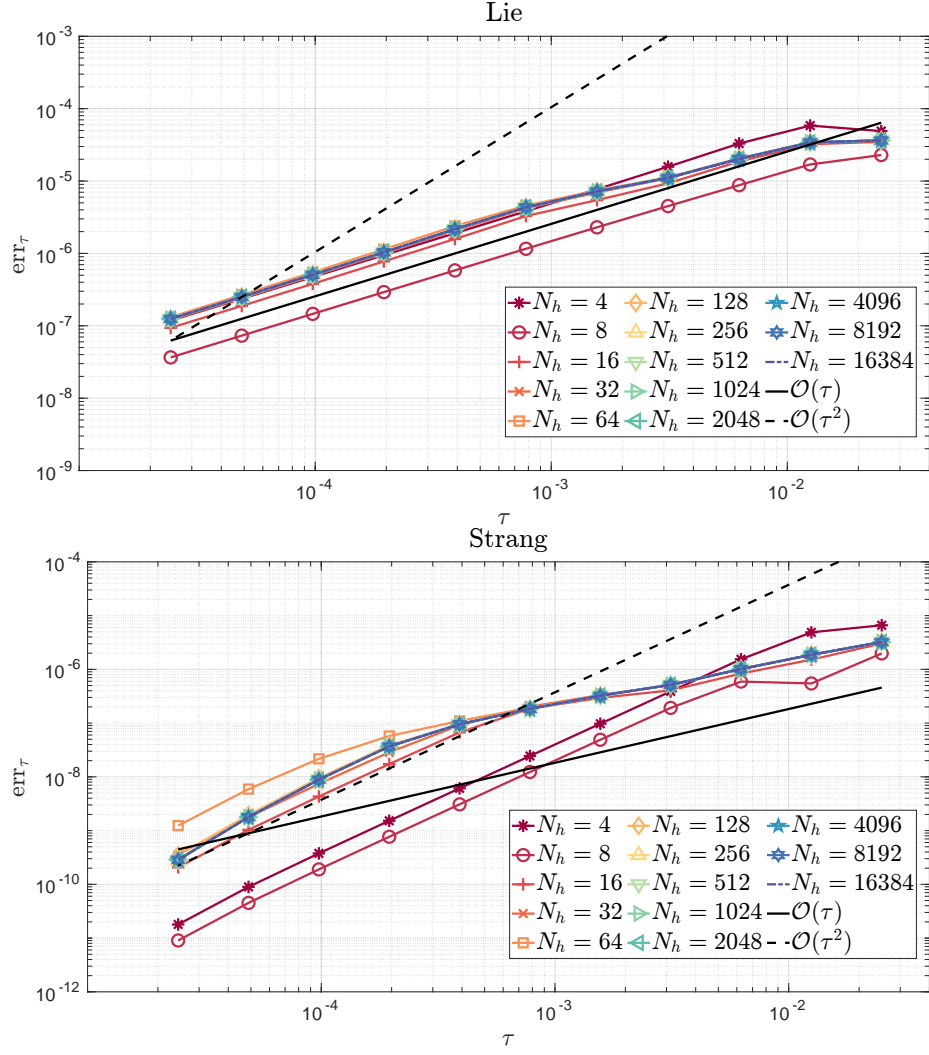
Figure 3: The errors err_τ for Lie (top) and Strang (bottom) splitting in a setting when Assumption 3 is fulfilled but Assumption 2 is not. Each curve corresponds to a different spatial discretization with N_h nodes.
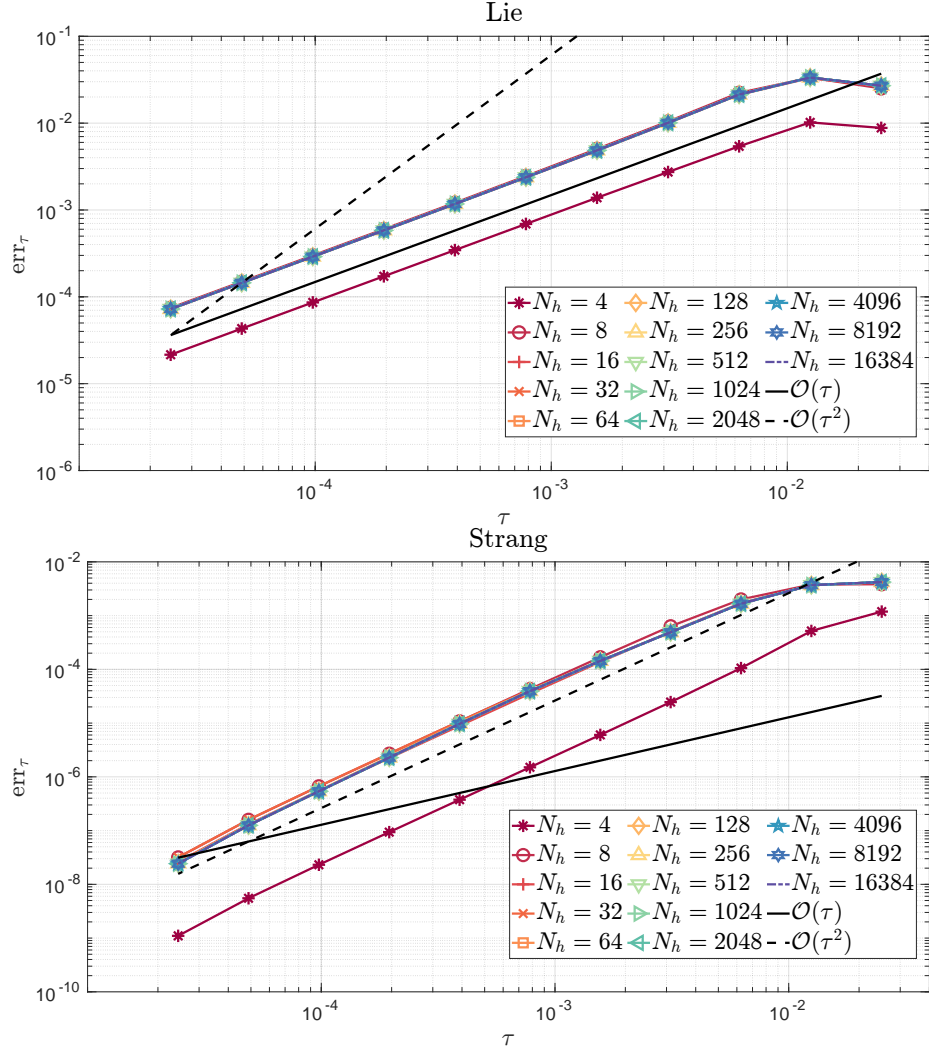
Figure 4: The errors err_τ for Lie (top) and Strang (bottom) splitting in a setting when Assumption 2 is fulfilled but Assumption 3 is not. Each curve corresponds to a different spatial discretization with N_h nodes.

5.4 Experiment 4

In this final experiment, none of the assumptions are fulfilled, and the errors shown in Figure 5 accordingly behave more erratically than in the other experiments. For Lie splitting, we observe convergence of order 1 for the coarsest spatial discretizations, but as N_h increases this decays and the errors increase. The effect is seen even more clearly for Strang splitting, where convergence of order 2 is initially observed, but for each increase of N_h, the time step size after which this is observed becomes smaller. We also see more clearly how the errors increase as N_h increases.

6 Conclusions

We have rigorously proved first- and second-order convergence of Lie and Strang splitting when applied to operator-valued DRE. These are the classical orders, which are to be expected for matrix-valued DREs under essentially no assumptions, since those are equivalent to systems of ODEs. However, as confirmed by the numerical experiments, if those matrix-valued equations are spatial discretizations of operator-valued equations then additional assumptions are necessary to avoid order reduction when the discretizations are refined. Our analysis relies on Assumption 2 or Assumption 3, either of which yield the requisite regularity of the exact solution. While we have not proven this, it seems unlikely that weaker assumptions would result in the full classical orders. However, we conjecture that assuming, e.g., $P_0 \in \mathcal{D}(A^r)$ with $0 < r < 1$ should lead to a solution $P \in C_p^r\big([0,T], \Sigma(H)\big)$ and thence to convergence with order r for Lie splitting. Investigating such modifications of the error analysis, as well as extensions to the case when B and E in the corresponding LQR problem are relatively unbounded operators, are topics for future work.

Acknowledgments

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.
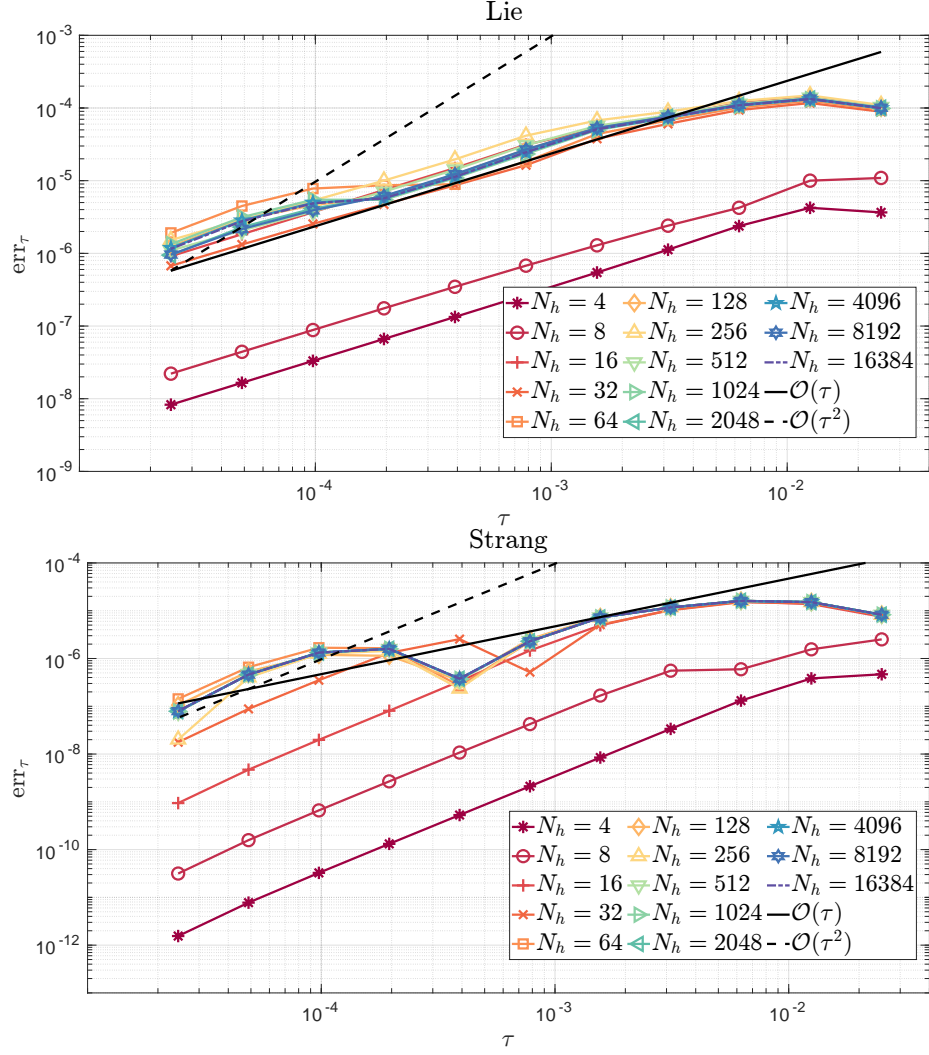
Figure 5: The errors err_τ for Lie (top) and Strang (bottom) splitting in a setting when neither Assumption 2 nor Assumption 3 is fulfilled. Each curve corresponds to a different spatial discretization with N_h nodes.

[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] 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.

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] G. J. Lord, C. E. Powell, and T. Shardlow. *An introduction to computational stochastic PDEs*. Cambridge Texts in Applied Mathematics. Cambridge University Press, New York, 2014. ISBN 978-0-521-72852-2. doi: 10.1017/CBO9781139017329.

[12] 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.

[13] 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.

[14] A. Ostermann, C. Piazzola, and H. Walach. 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.

[15] A. Pazy. *Semigroups of linear operators and applications to partial differential equations.* Applied mathematical sciences: 44. Springer, 1983. ISBN 0387908455.

[16] 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.

[17] 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.

[18] 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.

[19] T. Stillfjord. Convergence analysis for the exponential Lie splitting scheme applied to the abstract differential Riccati equation. Technical report, Lund University preprints, 2015.

[20] 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.