The Cayley transform as a time discretisation scheme*

V. Havu† and J. Malinen‡
Institute of Mathematics
Helsinki University of Technology
P.O.Box 1100,
FIN-02015 HUT, Finland

May 2, 2007

Abstract

We interpret the Cayley transform of linear (finite- or infinite-dimensional) state space systems as a numerical integration scheme of Crank–Nicolson type. The scheme is known as *Tustin's method* in the engineering literature, and it has the following important Hamiltonian integrator property: if Tustin's method is applied to a conservative (continuous time) linear system, then the resulting (discrete time) linear system is conservative in the discrete time sense. The purpose of this paper is to study the convergence of this integration scheme from the input/output point of view.

Keywords. Crank–Nicolson scheme, Cayley–Tustin transform, conservative system.
AMS Classification. 47A48, 65J10, 93C25, (34G10, 47N70, 65L70).

*This work is supported by the European Commission's 5th Framework Programme: Smart Systems; New materials, adaptive systems and their nonlinearities, HPRN-CT-2002-00284.

†Ville.Havu@tkk.fi
‡Jarmo.Malinen@tkk.fi

1 Introduction and motivation

This paper consists of two parts that can be read almost independently from each other. The first "system theory part" takes all of Section 1. It serves as a motivation for the second "numerical analysis part" that consists of Sections 2 – 5. All the new results are presented there, such as Theorems 1 and 2.

In Section 1 we discuss how time discretisation (1.2) of linear dynamical systems is related to the Cayley transform (understood in the sense of linear system theory). In finite-dimensional case, our dynamical systems are described by (1.1) but it is necessary to use the more general formulation (1.11) in infinite dimensions. Even the Cayley transform has to be generalized as explained in Section 1.3.

By Proposition 2, integration scheme (1.2) has the following nice property:

> If the original continuous time dynamical system (1.1) is conservative (as defined in Section 1.2), then the resulting discrete time system (1.4) satisfies an analogous energy equality.

Motivated by this observation, the convergence of a generalized, infinite-dimensional version of scheme (1.2) is investigated in the second part of the paper. The resulting numerical method can be used for input/output-simulation of input/output stable linear dynamical systems that are governed by PDE's from physics and engineering. Some of our results have been presented in [21] in a shortened form.

The real axis is denoted by $\mathbb{R}$ and the complex plane by $\mathbb{C}$, and we write $\mathbb{R}_+ = (0,\infty)$, $i\mathbb{R} = \{z : \operatorname{Re} z = 0\}$, $\mathbb{C}_+ = \{z : \operatorname{Re} z > 0\}$, and $\mathbb{D} = \{z : |z| < 1\}$. The usual Hardy spaces of X-valued analytic functions are denoted by $H^2(\mathbb{D};X)$, $H^\infty(\mathbb{D};X)$, $H^2(\mathbb{C}_+;X)$, and $H^\infty(\mathbb{C}_+;X)$ where X is a Banach space. By $C([0,\infty);X)$ we denote the X-valued norm-continuous functions on $[0,\infty)$, and the subset of compactly supported functions is $C_c([0,\infty);X)$. The space $C^n([0,\infty);X)$ denotes n times continuously differentiable functions for $n = 1, 2, \ldots$ where the derivatives at the endpoint is one-sided. If $X = \mathbb{C}$ above, then $\mathbb{C}$ is not written out explicitly. For $I \subset \mathbb{R}$, the Sobolev space $H^1(I)$ consists of complex-valued functions whose distribution derivative is in $L^2(I)$ – the set of square integrable functions. Bounded linear operators are denoted by $\mathcal{L}(X;Z)$ and $\mathcal{L}(X)$. Rest of the notation is either standard or introduced when used for the first time.

1.1 Cayley transform as Tustin time discretisation

For simplicity, we consider first the classical finite-dimensional case. Then the system S is described by the dynamical equations

$$S : \begin{cases} x'(t) = Ax(t) + Bu(t), \\ y(t) = Cx(t) + Du(t), \quad t \geq 0, \\ x(0) = x_0, \end{cases} \tag{1.1}$$

where $A \in \mathbb{C}^{n\times n}$, $B \in \mathbb{C}^{n\times m}$, $C \in \mathbb{C}^{p\times n}$, and $D \in \mathbb{C}^{p\times m}$. The input and the output of S are the signals $u(\cdot)$ and $y(\cdot)$, respectively. The function $x(\cdot)$ is called the state trajectory. Given a *discretisation parameter* $h > 0$, a slightly non-standard time discretisation of (1.1) of Crank–Nicolson type is given by

$$\begin{cases} \frac{x(jh)-x((j-1)h)}{h} & \approx A\frac{x(jh)+x((j-1)h)}{2} + Bu(jh), \\ y(jh) & \approx C\frac{x(jh)+x((j-1)h)}{2} + Du(jh), \quad j \geq 1 \\ x(0) & = x_0. \end{cases} \tag{1.2}$$

In engineering literature, this is sometimes called the *Tustin discretisation* of (1.1). Rewriting (1.2) gives the discrete time dynamics

$$\begin{cases} \frac{x_j^{(h)}-x_{j-1}^{(h)}}{h} & = A\frac{x_j^{(h)}+x_{j-1}^{(h)}}{2} + B\frac{u_j^{(h)}}{\sqrt{h}}, \\ \frac{y_j^{(h)}}{\sqrt{h}} & = C\frac{x_j^{(h)}+x_{j-1}^{(h)}}{2} + D\frac{u_j^{(h)}}{\sqrt{h}}, \quad j \geq 1, \\ x_0^{(h)} & = x_0, \end{cases} \tag{1.3}$$

where $u_j^{(h)}/\sqrt{h}$ is an approximation to $u(jh)$. The purpose of this paper is to characterize the convergence[1] of $y_j^{(h)}/\sqrt{h}$ to $y(jh)$ as $h \to 0$ in several different ways and under rather general assumptions.

Let us proceed to describe the connection of (1.1) – (1.3) to the *Cayley transform* in system theory. After some computations, equations (1.3) take the form

$$\phi_\sigma : \begin{cases} x_j^{(h)} & = \mathbf{A}_\sigma x_{j-1}^{(h)} + \mathbf{B}_\sigma u_j^{(h)}, \\ y_j^{(h)} & = \mathbf{C}_\sigma x_{j-1}^{(h)} + \mathbf{D}_\sigma u_j^{(h)}, \quad j \geq 1, \\ x_0^{(h)} & = x_0, \end{cases} \tag{1.4}$$

where $\sigma := 2/h$, and the operators $\mathbf{A}_\sigma$, $\mathbf{B}_\sigma$, $\mathbf{C}_\sigma$ and $\mathbf{D}_\sigma$ comprise the *discrete time linear system* (henceforth, DLS)

$$\phi_\sigma \equiv \begin{bmatrix} \mathbf{A}_\sigma & \mathbf{B}_\sigma \\ \mathbf{C}_\sigma & \mathbf{D}_\sigma \end{bmatrix} = \begin{bmatrix} (\sigma + A)(\sigma - A)^{-1} & \sqrt{2\sigma}(\sigma - A)^{-1}B \\ \sqrt{2\sigma}C(\sigma - A)^{-1} & \widehat{\mathcal{G}}(\sigma) \end{bmatrix}. \tag{1.5}$$

[1] To state this claim rigorously, we should define the sampling and interpolating operators $T_{2/h}$ and $T_{2/h}^*$. This is postponed to Section 2.2. Also note that we do not consider the approximation of $x(\cdot)$ in this paper but we restrict ourselves to the input/ouput framework.

Here $\widehat{\mathcal{G}}(\cdot)$ denotes the transfer function of system $S = \left[\begin{smallmatrix} A & B \\ C & D \end{smallmatrix}\right]$ in (1.1), and it is defined by $\widehat{\mathcal{G}}(s) = D + C(s-A)^{-1}B$ for all $s \in \rho(A)$. Then the transfer function $\widehat{\mathcal{D}}_\sigma(\cdot)$ of ϕ_σ clearly satisfies

$$\widehat{\mathcal{D}}_\sigma(z) := \mathbf{D}_\sigma + z\mathbf{C}_\sigma(I - z\mathbf{A}_\sigma)^{-1}\mathbf{B}_\sigma = \widehat{\mathcal{G}}\left(\frac{1-z}{1+z}\sigma\right) \tag{1.6}$$

for all $z^{-1} \in \rho(\mathbf{A}_\sigma)$. The mapping $S \mapsto \phi_\sigma$ described above is called the Cayley transform of continuous time systems to discrete time systems. The purpose of this paper is to show that (1.2) successfully approximates (1.1) in a context of input/output mappings of infinite-dimensional linear dynamical systems. Hence, the DLS ϕ_σ can be regarded as a convergent time discretisation of S.

Out of our convergence results, Proposition 4 and Lemma 1 are stated in the frequency domain. Lemma 1 provides a speed estimate for the convergence that is uniform on the compact subsets of frequencies; see also Corollary 1 for a more intuitive but less sharp estimate. As a consequence of Lemma 1, more practical Theorems 1 and 2 are given in time domain but unfortunately without a speed estimate. It is finally shown that Theorem 2 cannot be improved by a speed estimate similar to Lemma 1.

1.2 Infinite-dimensional linear systems

Even though we considered above only matrix systems (1.1), the Cayley transform can be defined similarly to (1.5) for any *system node* S. System nodes are a functional analytic framework for presenting linear dynamical systems with possibly infinite-dimensional state spaces – including boundary control systems defined by PDE's. System nodes are discussed in, e.g., Malinen, Staffans and Weiss [25] but we review the construction below[2].

Let X be a Hilbert space and let $A\colon \mathrm{dom}\,(A) \subset X \to X$ be a closed, densely defined linear operator with a nonempty resolvent set $\rho(A)$. Take $\alpha \in \rho(A)$, and define $\|x\|_{X_1} = \|(\alpha - A)x\|_X$ for each $x \in \mathrm{dom}\,(A)$. Then $\|\cdot\|_{X_1}$ is a norm on $\mathrm{dom}\,(A)$ which makes it into a Hilbert space called X_1. It follows that $A \in \mathcal{L}(X_1; X)$. The space X_{-1} is defined as the completion of X with respect to the norm $\|x\|_{X_{-1}} = \|(\alpha - A)^{-1}x\|_X$ which makes X_{-1} a Hilbert space. We have now constructed a triple of Hilbert spaces $X_1 \subset X \subset X_{-1}$ with dense and continuous embeddings — the *rigged Hilbert spaces* induced by A and X. A different choice of $\alpha \in \rho(A)$ leads to equivalent norms in X_1 and X_{-1} but it does not change the spaces themselves. The operator A

[2]The rest of this section serves only as a motivation and background. An already well-motivated reader may skip to Section 2 without any loss to read the rest of this paper.

has a unique extension (by density and continuity) to an operator $A_{-1} \in \mathcal{L}(X; X_{-1})$, known as the *Yosida extension* of A.

Definition 1. *Let U, X and Y be Hilbert spaces*[3]*. An operator*

$$S := \begin{bmatrix} A\&B \\ C\&D \end{bmatrix} : \begin{bmatrix} X \\ U \end{bmatrix} \supset \operatorname{dom}(S) \to \begin{bmatrix} X \\ Y \end{bmatrix}$$

is called a system node on (U, X, Y) if it has the following structure:

(i) *A is a generator of a strongly continuous semigroup on X with its Yosida extension $A_{-1} \in \mathcal{L}(X; X_{-1})$ as explained above.*

(ii) $B \in \mathcal{L}(U; X_{-1})$.

(iii) $\operatorname{dom}(S) := \left\{ \left[\begin{smallmatrix} x \\ u \end{smallmatrix}\right] \in \left[\begin{smallmatrix} X \\ U \end{smallmatrix}\right] \;\middle|\; A_{-1}x + Bu \in X \right\}$.

(iv) $A\&B = \begin{bmatrix} A_{-1} & B \end{bmatrix}_{|\operatorname{dom}(S)}$.

(v) $C\&D \in \mathcal{L}(\operatorname{dom}(S); Y)$*; we use on* $\operatorname{dom}(S)$ *the graph norm of $A\&B$:*

$$\left\| \left[\begin{smallmatrix} x \\ u \end{smallmatrix}\right] \right\|^2_{\operatorname{dom}(S)} := \|x\|^2_X + \|u\|^2_U + \|A_{-1}x + Bu\|^2_X.$$

Let now $S = \left[\begin{smallmatrix} A\&B \\ C\&D \end{smallmatrix}\right]$ be a system node on Hilbert spaces (U, X, Y) as in Definition 1. We call $A \in \mathcal{L}(X_1; X)$ the *main operator* or *semigroup generator* of S, $B \in \mathcal{L}(U; X_{-1})$ is its *control operator*, and $C\&D \in \mathcal{L}(\operatorname{dom}(S); Y)$ is its *combined observation/feedthrough operator*. From the last operator we can extract $C \in \mathcal{L}(X_1; Y)$, the *observation operator* of S, defined by

$$Cx := C\&D \begin{bmatrix} x \\ 0 \end{bmatrix}, \qquad x \in X_1. \tag{1.7}$$

It is trivial that $A\&B \in \mathcal{L}(\operatorname{dom}(S), X)$. A short computation shows that for each $\alpha \in \rho(A)$, the operator $E_\alpha := \left[\begin{smallmatrix} I & (\alpha - A_{-1})^{-1}B \\ 0 & I \end{smallmatrix}\right]$ is a bounded bijection from $\left[\begin{smallmatrix} X \\ U \end{smallmatrix}\right]$ onto itself and also from $\left[\begin{smallmatrix} X_1 \\ U \end{smallmatrix}\right]$ onto $\operatorname{dom}(S)$. Since $\left[\begin{smallmatrix} X_1 \\ U \end{smallmatrix}\right]$ is dense in $\left[\begin{smallmatrix} X \\ U \end{smallmatrix}\right]$, this implies that $\operatorname{dom}(S)$ is dense in $\left[\begin{smallmatrix} X \\ U \end{smallmatrix}\right]$, too. It takes more reasoning to see that S, in fact, is closed as a densely defined operator from $\left[\begin{smallmatrix} X \\ U \end{smallmatrix}\right]$ to $\left[\begin{smallmatrix} X \\ Y \end{smallmatrix}\right]$. Since the second column of E_α maps U into $\operatorname{dom}(S)$, we can define the *transfer function* of S by

$$\widehat{\mathcal{G}}(s) := C\&D \begin{bmatrix} (s - A_{-1})^{-1}B \\ I \end{bmatrix}, \qquad s \in \rho(A), \tag{1.8}$$

[3] We shall use the notation $\left[\begin{smallmatrix} X \\ Y \end{smallmatrix}\right]$ for $X \times Y$.

which is an $\mathcal{L}(U;Y)$-valued analytic function. A system node is called *input/output* or *I/O stable* if $\mathbb{C}_+ \subset \rho(A)$ and $\widehat{\mathcal{G}}(\cdot) \in H^\infty(\mathbb{C}_+;\mathcal{L}(U,Y))$.

In above construction, the operator node S, the observation operator C, and the transfer function $\widehat{\mathcal{G}}$ are determined by the operators A, B and $C\&D$. Alternatively, S and $\widehat{\mathcal{G}}$ may be constructed from A, B, C and the value $\widehat{\mathcal{G}}(\alpha)$ at one point in $\alpha \in \rho(A)$; see [25, Section 2] for details.

Example 1. *For any $m, n, p \in \mathbb{N}$, take any matrices $A \in \mathbb{C}^{n\times n}$, $B \in \mathbb{C}^{n\times m}$, $C \in \mathbb{C}^{p\times n}$, and $D \in \mathbb{C}^{p\times m}$ as in Section 1.1. Then the block matrix $S' := \left[\begin{smallmatrix} A & B \\ C & D \end{smallmatrix}\right]$ is a system node on $(\mathbb{C}^m, \mathbb{C}^n, \mathbb{C}^p)$ with $\operatorname{dom}(S') = \left[\begin{smallmatrix} \mathbb{C}^n \\ \mathbb{C}^m \end{smallmatrix}\right]$, $A_1 = A = A_{-1}$, $A\&B = \begin{bmatrix} A & B \end{bmatrix}$, and $C\&D = \begin{bmatrix} C & D \end{bmatrix}$. Also (1.8) is equivalent with $\widehat{\mathcal{G}}(s) = C(s-A)^{-1}B + D$ for all $s \in \rho(A)$.*

In Example 1, we have $D = \lim_{|s|\to\infty} \widehat{\mathcal{G}}(s)$. Such an operator D is called the *feedthrough operator* of $S = \left[\begin{smallmatrix} A\&B \\ C\&D \end{smallmatrix}\right]$ whenever the defining limit exists in some operator topology. We remark that not all system nodes satisfying $\dim X = \infty$ have a well-defined feedthrough operator, and this is the reason why we use the combined operator $C\&D$ in Definition 1. System nodes known as *regular well-posed systems* possess feedthrough operators; see, e.g., Staffans and Weiss [34, 35], and Weiss [38].

The main reason for defining system nodes is that the "finite-dimensional" dynamical equations (1.1) can be generalized for any system nodes. Indeed, there exists a unique $x \in C^1([0,\infty);X)$ such that

$$\begin{cases} x'(t) = A_{-1}x(t) + Bu(t), \quad t \geq 0, \\ x(0) = x_0 \end{cases} \tag{1.9}$$

holds for any input $u \in C^2([0,\infty);U)$ and any initial state $x_0 \in X$ for which the compatibility condition $\left[\begin{smallmatrix} x_0 \\ u(0) \end{smallmatrix}\right] \in \operatorname{dom}(S)$ holds. Moreover, $\left[\begin{smallmatrix} x(\cdot) \\ u(\cdot) \end{smallmatrix}\right] \in C([0,\infty);\operatorname{dom}(S))$ and because $C\&D \in \mathcal{L}(\operatorname{dom}(S);U)$, the output signal given by

$$y(t) = C\&D \begin{bmatrix} x(t) \\ u(t) \end{bmatrix} \tag{1.10}$$

is well-defined and continuous for all $t \geq 0$. We may write (1.9) and (1.10) shortly as

$$\begin{bmatrix} \dot{x}(t) \\ y(t) \end{bmatrix} = S \begin{bmatrix} x(t) \\ u(t) \end{bmatrix}, \qquad t \geq 0, \qquad x(0) = x_0, \tag{1.11}$$

which is the required generalization of (1.1) to system node S.

The role of the transfer function (1.8) is the same as in the finite dimensional case. Indeed, define the Laplace-transform as usual by

$$\hat{f}(s) \equiv (\mathcal{L}f)(s) = \int_0^\infty e^{-st} f(t)\, dt \quad \text{for all} \quad s \in \mathbb{C}_+. \tag{1.12}$$

Then $\hat{y}(s) = \widehat{\mathcal{G}}(s)\hat{u}(s)$ for all $s \in \mathbb{C}_+$ with the estimate

$$\|y\|_{L^2(\mathbb{R}_+;Y)} \leq \sup_{s\in\mathbb{C}_+} \|\widehat{\mathcal{G}}(s)\|_{\mathcal{L}(U;Y)} \|u\|_{L^2(\mathbb{R}_+;U)} \tag{1.13}$$

if $u(\cdot)$ and $y(\cdot)$ are related by (1.11) with $x_0 = 0$ (and the integral in (1.12) converges). This mapping $u(\cdot) \mapsto y(\cdot)$ (with $x_0 = 0$) is called the *input/output mapping* of S. It has by density a unique extension to a bounded operator from $L^2(\mathbb{R}_+;U)$ into $L^2(\mathbb{R}_+;Y)$ assuming that S is I/O stable. These and many other facts can be found in [25, Section 2] with all details.

1.3 Cayley–Tustin transform in infinite dimensions

We now describe how the Cayley transform can be extended to system nodes S with infinite-dimensional state spaces. The Cayley transform $\phi_\sigma \equiv \left[\begin{smallmatrix} \mathbf{A}_\sigma & \mathbf{B}_\sigma \\ \mathbf{C}_\sigma & \mathbf{D}_\sigma \end{smallmatrix}\right]$ of S is simply the DLS defined by

$$\phi_\sigma := \begin{bmatrix} (\sigma + A)(\sigma - A)^{-1} & \sqrt{2\sigma}(\sigma - A_{-1})^{-1}B \\ \sqrt{2\sigma}C(\sigma - A)^{-1} & \widehat{\mathcal{G}}(\sigma) \end{bmatrix} \tag{1.14}$$

for any $\sigma \in \rho(A) \cap \mathbb{R}_+$. When comparing to the matrix formula (1.5), we see that A has been replaced by its extension A_{-1} in one place. The observation operator C and the transfer function $\widehat{\mathcal{G}}(\cdot)$ are now defined through (1.7) and (1.8), respectively. The transfer function of $\widehat{\mathcal{D}}_\sigma(\cdot)$ of ϕ_σ — together with its relation to $\widehat{\mathcal{G}}(\cdot)$ — is described by (1.6) without change.

Proposition 1. *Let $\sigma > 0$ and S be a system node whose main operator satisfies $\mathbb{R}_+ \subset \rho(A)$. Then S is (continuous time) I/O stable if and only if its Cayley transform ϕ_σ is (discrete time) I/O stable.*

This follows by applying the spectral mapping theorem to the identity $\mathbf{A}_\sigma = (\sigma + A)(\sigma - A)^{-1}$, using (1.6), and recalling that the DLS ϕ_σ is I/O -stable if and only if $\sigma(\mathbf{A}_\sigma) \subset \overline{\mathbb{D}}$ and $\widehat{\mathcal{D}}_\sigma(\cdot) \in H^\infty(\mathbb{D};\mathcal{L}(U;Y))$.

From now on we shall not use equations (1.1) – (1.3) and (1.5) (which were given only as an introduction) any longer but their infinite-dimensional generalized versions (1.9) – (1.11) and (1.14) instead. The approximating trajectories will be given by (1.4) even in the general case, defining the required operators by (1.14) and the identity $\phi_\sigma \equiv \left[\begin{smallmatrix} \mathbf{A}_\sigma & \mathbf{B}_\sigma \\ \mathbf{C}_\sigma & \mathbf{D}_\sigma \end{smallmatrix}\right]$.

There exists an extensive general literature on the Cayley transform of systems but we shall not make an account of it; see, e.g., Ober and Montgomery-Smith [28] and the numerous other references given in [33]. The idea of using the Cayley transform for the simulation of linear systems is not

new, either. In finite dimensions, the method described by (1.3) was already discovered in 1940's by Tustin, and it is known as the *Tustin transform* in digital and sampled-data control circles; see, e.g., [29, p. 137].

The Cayley transform can be used in numerical analysis in a way that is completely different from Tustin's approach; see Arov and Gavrilyuk [1], Gavrilyuk [9, 10, 11], and Gavrilyuk and Makarov [12, 13, 14, 15, 16, 17, 18]. The analytical and numerical solution of differential equations of type $x^{(n)} = Lx$ and $x^{(n)} = Lx + f$ for $n = 1, 2$, is considered with various assumptions on operator L that are relevant either in Hilbert or in Banach space context. The numerical method proposed by these authors is *spectral* in the sense that the discretisation is a truncation in the Laguerre polynomial basis. This is in contrast to Tustin's approach which is a time-domain *difference approximation* instead.

1.4 Tustin's discretisation preserves conservativity

The system node S is *(scattering) energy preserving* if for all $T > 0$ the energy balance equation

$$\|x(T)\|_X^2 + \int_0^T \|y(t)\|_Y^2 dt = \|x_0\|_X^2 + \int_0^T \|u(t)\|_U^2 dt \tag{1.15}$$

holds, where u, x, y and x_0 are as in (1.9) — (1.11). For any energy preserving S, the main operator A is maximally dissipative and $\mathbb{C}_+ \subset \rho(A)$. Then equation (1.14) defines the Cayley transform ϕ_σ for all $\sigma > 0$. Letting $T \to \infty$ in (1.15) shows that the input/output mapping of an energy preserving S is a contraction from $L^2(\mathbb{R}_+; U)$ into $L^2(\mathbb{R}_+; Y)$, and hence its transfer function satisfies $\|\widehat{\mathcal{G}}(s)\|_{\mathcal{L}(U;Y)} \leq 1$ for all $s \in \mathbb{C}_+$.

If both $S = \left[\begin{smallmatrix} A\&B \\ C\&D \end{smallmatrix}\right]$ and its *dual node* $S^d = \left[\begin{smallmatrix} [A\&B]^d \\ [C\&D]^d \end{smallmatrix}\right]$ are scattering energy preserving, then $\left[\begin{smallmatrix} A\&B \\ C\&D \end{smallmatrix}\right]$ is called *(scattering) conservative*. The dual node S^d is defined simply as the unbounded adjoint of S when it is regarded as a closed, densely defined operator from $\left[\begin{smallmatrix} X \\ U \end{smallmatrix}\right]$ to $\left[\begin{smallmatrix} X \\ Y \end{smallmatrix}\right]$ (see the discussion following (1.7)). We remark that it is now a nontrivial fact that the adjoint of S actually is a system node in the sense of Definition 1. For details, we refer to [25, Proposition 2.4, and Definitions 3.1 and 4.1].

We say that the DLS $\phi = \left[\begin{smallmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{smallmatrix}\right]$ is *energy preserving* if the block matrix $\left[\begin{smallmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{smallmatrix}\right]$ is isometric from $\left[\begin{smallmatrix} X \\ U \end{smallmatrix}\right]$ into $\left[\begin{smallmatrix} X \\ Y \end{smallmatrix}\right]$. Then, and only then, the discrete time balance equation

$$\|x_N\|_X^2 - \|x_0\|_X^2 = \sum_{j=1}^{N} \|u_{j-1}\|_U^2 - \sum_{j=1}^{N} \|y_{j-1}\|_Y^2$$

is satisfied for all $N \geq 1$, all initial values $x_0 \in X$ and all sequences $\{u_j\}$, $\{x_j\}$ and $\{y_j\}$ satisfying

$$\begin{cases} x_{j+1} &= \mathbf{A}x_j + \mathbf{B}u_j, \\ y_{j+1} &= \mathbf{C}x_j + \mathbf{D}u_j, \quad j \geq 0. \end{cases}$$

The DLS ϕ is *conservative* if both ϕ and the dual DLS $\phi^d := \left[\begin{smallmatrix} \mathbf{A}^* & \mathbf{C}^* \\ \mathbf{B}^* & \mathbf{D}^* \end{smallmatrix}\right]$ (defined as the adjoint of a bounded block operator) are energy preserving. If the spaces U and Y coincide, then ϕ is conservative if and only if the block operator $\left[\begin{smallmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{smallmatrix}\right]$ is unitary on $\left[\begin{smallmatrix} X \\ U \end{smallmatrix}\right]$. For the proof of the next Proposition, see [25, Theorem 3.2(v) and Theorem 4.2(iii)]:

Proposition 2. *The Cayley transform ϕ_σ of an energy preserving system node S is an energy preserving DLS. Moreover, such ϕ_σ is (discrete time) conservative if and only if S is a conservative.*

The reason for preferring the discretisation by (1.4) and (1.14) for energy preserving and conservative problems (1.11) is due to Proposition 2. We emphasize that Proposition 4, Lemma 1, and Theorem 2 below let us conclude that (1.4) and (1.14) can be interpreted as a convergent time discretisation scheme for *all I/O stable* – including many non-conservative – system nodes satisfying $\dim U = \dim Y = 1$.

This is easy to understand because our results of are formulated in terms of transfer functions and input/output mappings, and hence they do not depend at all on the particular choice of the state space realization of type (1.11). The only connection to system nodes is via the Cayley transform (1.6) between continuous and discrete time transfer functions.

Conservative system nodes are known in operator theory as *operator colligations* or *Livšic – Brodskiĭ nodes*. Much classical literature exists for them, see, e.g., Arov and Nudelman [2], Ball and Staffans [3], Brodskiĭ [5, 7, 6], Livšic [23], Livšic and Yantsevich [22], Sz.-Nagy and Foiaş [36], Smuljan [30], and Staffans [31, 32, 33]. Operator theory techniques for proving conservativity in applications are given in Malinen, Staffans and Weiss [25], and Tucsnak and Weiss [39, 37]. The special case of boundary control systems is further studied in Malinen [24], and Malinen and Staffans [26, 27]; see also Gorbachuk and Gorbachuk [19] and the references therein.

In numerical analysis, integration schemes that preserve energy equalities or more complex invariants of the system are called *Hamiltonian* or *symplectic*, respectively. The Cranck-Nicolson scheme (1.3) for linear systems is a lowest order symplectic integration scheme from the family of Gauss quadrature based Runge-Kutta methods. There exists an extensive literature of symplectic schemes; see, e.g., Hairer, Lubich and Wanner [20].

2 Approximation of the input/output mapping

In this section, we rewrite the discretisation (1.4) of the infinite-dimensional dynamical system (1.11) in operator theory language. After that we explain how its convergence can be studied as an approximation of the Laplace transform.

From now on we make it a standing assumption that $\widehat{\mathcal{G}}(\cdot)$ is a (possibly non-rational) transfer function of an I/O stable system node with scalar input and output spaces $U = Y = \mathbb{C}$. This means that $\widehat{\mathcal{G}}(\cdot) \in H^\infty(\mathbb{C}_+)$ or, equivalently, $\widehat{\mathcal{D}}_\sigma(\cdot)$ given by (1.6) satisfies $\widehat{\mathcal{D}}_\sigma(\cdot) \in H^\infty(\mathbb{D})$; see Proposition 1.

2.1 Spaces, norms and transforms

We use the norm

$$\|f\|^2_{H^2(\mathbb{C}_+)} = \frac{1}{2\pi} \sup_{x>0} \int_{-\infty}^{\infty} |f(x+yi)|^2 \, dy$$

for the Hardy space $H^2(\mathbb{C}_+)$. Then the Laplace transform is defined by (1.12) is unitary from $L^2(\mathbb{R}_+)$ onto $H^2(\mathbb{C}_+)$. The norm of $H^2(\mathbb{D})$ is given by $\|\phi\|^2_{H^2(\mathbb{D})} = \sum_{j\geq 0} |\phi_j|^2$ for $\phi(z) = \sum_{j\geq 0} \phi_j z^j$, and it makes the Z-transform unitary from $\ell^2(\mathbb{Z}_+) \to H^2(\mathbb{D})$. If, say, $f \in C_c(\mathbb{R})$ in (1.12), then $(\mathcal{L}f)(s)$ is well defined for all $s \in i\mathbb{R}$, too. The function $i\omega \mapsto (\mathcal{L}f)(i\omega)$ is then the Fourier transform of f.

By $\widehat{\mathcal{D}}_\sigma : H^2(\mathbb{D}) \to H^2(\mathbb{D})$ denote the multiplication operator satisfying $(\widehat{\mathcal{D}}_\sigma \tilde{u})(z) = \widehat{\mathcal{D}}_\sigma(z)\tilde{u}(z)$ for all $z \in \mathbb{D}$ and $\sigma > 0$. Similarly, denote by $\widehat{\mathcal{G}} : H^2(\mathbb{C}_+) \to H^2(\mathbb{C}_+)$ the multiplication operator satisfying $(\widehat{\mathcal{G}}\hat{u})(s) = \widehat{\mathcal{G}}(s)\hat{u}(s)$ for all $s \in \mathbb{C}_+$. The operators $\widehat{\mathcal{D}}_\sigma$ and $\widehat{\mathcal{G}}$ are unitarily equivalent to the input/output mappings of ϕ_σ and S, respectively. The correspondence (1.6) takes the form of the similarity transform

$$\widehat{\mathcal{G}} = \mathcal{C}_\sigma^{-1} \widehat{\mathcal{D}}_\sigma \mathcal{C}_\sigma, \tag{2.1}$$

where the *composition operator* is defined by $(\mathcal{C}_\sigma F)(z) := F(\frac{1-z}{1-z}\sigma)$ for all $z \in \mathbb{D}$ and $F : \mathbb{C}_+ \to \mathbb{C}$. It is easy to see that $(\mathcal{C}_\sigma^{-1} f)(s) := f(\frac{s-\sigma}{s+\sigma})$ for all $s \in \mathbb{C}_+$ and all $f : \mathbb{D} \to \mathbb{C}$. Hence we have $\mathcal{M}_\sigma \mathcal{C}_\sigma^{-1} f = F$ where $F(s) = \frac{\sqrt{2/\sigma}}{1+s/\sigma} f(\frac{s-\sigma}{s+\sigma})$ and $\mathcal{M}_\sigma$ denotes the multiplication operator by the function $s \mapsto \frac{\sqrt{2/\sigma}}{1+s/\sigma}$.

Proposition 3. *The operator $\mathcal{M}_\sigma \mathcal{C}_\sigma^{-1} : H^2(\mathbb{D}) \to H^2(\mathbb{C}_+)$ is unitary.*

This holds because the sequence $\left\{ \frac{\sqrt{2/\sigma}}{1+s/\sigma} \left(\frac{s-\sigma}{s+\sigma}\right)^j \right\}_{j\geq 0}$ is an orthonormal basis for $H^2(\mathbb{C}_+)$ for each $\sigma > 0$.

2.2 Discretising operators

By T_σ we denote a discretising (or sampling) bounded linear operator $T_\sigma : L^2(\mathbb{R}_+) \to H^2(\mathbb{D})$. The adjoint T_σ^* of T_σ maps then $H^2(\mathbb{D}) \to L^2(\mathbb{R}_+)$, and it is typically an interpolating operator. The operator T_σ can be defined in many ways but in this paper we use the mean value sampling

$$(T_\sigma u)(z) = \sum_{j\geq 1} u_j^{(h)} z^j \quad \text{where} \quad \frac{u_j^{(h)}}{\sqrt{h}} = \frac{1}{h}\int_{(j-1)h}^{jh} u(t)\,dt \tag{2.2}$$

with $h = 2/\sigma$ (recall (1.3) and (1.4)). Then the adjoint T_σ^* is given by

$$\left(T_\sigma^*\tilde{v}\right)(t) = \frac{1}{\sqrt{h}}\sum_{j\geq 1} v_j \chi_{[(j-1)h,jh]}(t), \tag{2.3}$$

where $\tilde{v}(z) = \sum_{j\geq 0} v_j z^j \in H^2(\mathbb{D})$ and $\chi_I(\cdot)$ denotes the characteristic function of the interval I. It is worth noticing that the operator T_σ is a coisometry, i.e., T_σ^* is an isometry:

$$\|T_\sigma^*\tilde{v}\|_{L^2(\mathbb{R}_+)}^2 = \frac{1}{h}\int_0^\infty |\sum_{j\geq 1} v_j \chi_{[(j-1)h,jh]}|^2\,dt = \frac{1}{h}\int_0^\infty \sum_{j\geq 1} |v_j|^2 \chi_{[(j-1)h,jh]}\,dt \tag{2.4}$$

$$= \frac{1}{h}\sum_{j\geq 1} |v_j|^2 \int_0^\infty \chi_{[(j-1)h,jh]}\,dt = \sum_{j\geq 1} |v_j|^2 = \|\tilde{v}\|_{H^2(\mathbb{D})}^2.$$

The operator T_σ itself is not isometric since $\ker(T_\sigma) \neq \{0\}$.

2.3 Approximation of the Laplace transform

Let us now use the discrete time trajectories of (1.4) to approximate the continuous time dynamics in (1.11) using the discretisation and sampling by operators T_σ and T_σ^*.

Let $u \in L^2(\mathbb{R}_+)$ and assume zero initial states for both the system (1.9) — (1.11) and its Tustin discretisation (1.4). The input signal of (1.4) is the discretised signal $T_\sigma u$. If we transform the output $\{y^{(h)}\}_{j\geq 0}$ of (1.4) into a continuous time signal by applying the interpolating operator T_σ^* to it, we obtain the signal $T_\sigma^*\widehat{\mathcal{D}}_\sigma T_\sigma u$. On the other hand, the output of the continuous time dynamics (1.11) is given by $\mathcal{L}^*\widehat{\mathcal{G}}\mathcal{L}u$. Our task is to show that at least for some nice $u \in L^2(\mathbb{R}_+)$ and $T > 0$, we have the convergence

$$\|T_\sigma^*\widehat{\mathcal{D}}_\sigma T_\sigma u - \mathcal{L}^*\widehat{\mathcal{G}}\mathcal{L}u\|_{L^2(0,T)} \to 0 \tag{2.5}$$

as $\sigma \to \infty$. This will be achieved in Theorem 2. By Proposition 3 and equation (2.1) we see that

$$\begin{aligned} T_\sigma^* \widehat{\mathcal{D}}_\sigma T_\sigma &= T_\sigma^* \left(\mathcal{C}_\sigma \mathcal{M}_\sigma^{-1}\right) \cdot \widehat{\mathcal{G}} \cdot \left(\mathcal{M}_\sigma \mathcal{C}_\sigma^{-1}\right) T_\sigma \\ &= T_\sigma^* \left(\mathcal{M}_\sigma \mathcal{C}_\sigma^{-1}\right)^{-1} \cdot \widehat{\mathcal{G}} \cdot \left(\mathcal{M}_\sigma \mathcal{C}_\sigma^{-1}\right) T_\sigma = \left(\mathcal{M}_\sigma \mathcal{C}_\sigma^{-1} T_\sigma\right)^* \cdot \widehat{\mathcal{G}} \cdot \left(\mathcal{M}_\sigma \mathcal{C}_\sigma^{-1} T_\sigma\right) \end{aligned}$$

since the multiplication operator $\mathcal{M}_\sigma$ commutes with $\widehat{\mathcal{G}}$. Motivated by this equation and by (2.5), we inquire whether the operators $L_\sigma := \mathcal{M}_\sigma \mathcal{C}_\sigma^{-1} T_\sigma$ are in some sense close[4] to the Laplace transform $\mathcal{L}$ when $\sigma \to \infty$. Thus, another aim of this paper is to give stronger versions of the following proposition:

Proposition 4. *For any $u \in C_c(\mathbb{R}_+)$ and $s \in \mathbb{C}_+$, we have*

$$(\mathcal{L}u)(s) = \lim_{\sigma \to \infty} (L_\sigma u)(s),$$

where L_σ is defined as above.

Proof. Defining T_σ by (2.2) we get

$$\begin{aligned} (L_\sigma u)(s) &= \frac{\sqrt{2/\sigma}}{1+s/\sigma} \sum_{j \geq 1} \left(\frac{1}{h} \int_{(j-1)h}^{jh} u(t)\, dt \right) \left(\frac{\sigma - s}{\sigma + s} \right)^j \qquad (2.6) \\ &= \frac{1}{1+s/\sigma} \sum_{j \geq 1} \left(\int_0^\infty \chi_{[(j-1)h, jh]}(t) \left(\frac{\sigma - s}{\sigma + s} \right)^j u(t)\, dt \right) \\ &= \int_0^\infty K_{s,\sigma}(t) u(t)\, dt, \end{aligned}$$

where $\sigma = 2/h$ and

$$K_{s,\sigma}(t) = \frac{1}{1+s/\sigma} \sum_{j \geq 1} \chi_{[(j-1)h, jh]}(t) \left(1 - \frac{2s}{s+\sigma} \right)^j. \qquad (2.7)$$

Now, if j is such that $t \in [(j-1)h, jh]$, then we obtain from the previous

$$K_{s,\sigma}(t) \approx \frac{1}{1+s/\sigma} \left(1 - \frac{s}{s/2 + \sigma/2} \right)^{(\sigma/2)\cdot t} \to e^{-st} \text{ as } \sigma \to \infty.$$

[4]Note that by Proposition 3 and equality (2.4), we see that each $L_\sigma : L^2(\mathbb{R}_+) \to H^2(\mathbb{C}_+)$ is a coisometry. The Laplace transform is a unitary mapping between the same spaces. Hence, the convergence of $L_\sigma \to \mathcal{L}$ must be rather weak.

We conclude that $\lim_{\sigma\to\infty} K_{s,\sigma}(t) = e^{-st}$ for all $s \in \mathbb{C}_+$ and $t \geq 0$. Moreover, for each fixed $s \in \mathbb{C}_+$ and $\sigma \geq 2|s|$ we have

$$\begin{aligned}|K_{s,\sigma}(t)| &\leq 2 \cdot \left(1 + \frac{2|s|}{\sigma - |s|}\right)^{(\sigma/2)\cdot t} \\ &\leq 2 \cdot \left(1 + \frac{2|s|}{\sigma - |s|}\right)^{(\sigma-|s|)t/2} \cdot \left(1 + \frac{2|s|}{\sigma - |s|}\right)^{|s|t/2} \leq 2\left(e\sqrt{3}\right)^{|s|t}.\end{aligned}$$

The proposition now follows from the Lebesgue dominated convergence theorem, as the integrand in (2.6) has a compact support. □

3 A pointwise convergence estimate

Our most important preliminary result Lemma 1 is given in this section. We obtain a uniform speed estimate for the convergence of $(L_\sigma u)(i\omega) \to (\mathcal{L}u)(i\omega)$ for $i\omega \in K$ where $K \subset i\mathbb{R}$ is compact.

Before that some new definitions and notations must be given: Let $I_j = ((j-1)h, jh] = (t_{j-1}, t_j]$ and $t_{j-1/2} = \frac{1}{2}(t_{j-1} + t_j)$. For $u \in L^2(\mathbb{R}_+)$, let $I_{h,s}u$ be the piecewise linear (with jumps) interpolating function, defined by

$$(I_{h,s}u)(t) = \bar{u}_{j,h} + \frac{c_j(h,s)}{h}(t - t_{j-1/2}), \quad t \in I_j, \tag{3.1}$$

where $\bar{u}_{j,h} = \frac{1}{h}\int_{I_j} u(t)\,dt$ and the defining sequence $\{c_j(h,s)\}_{j\geq 1}$ (depending on two parameters h and s) will be later chosen in a particular way. Let P_h denote the orthogonal projection in $L^2(\mathbb{R}_+)$ onto the subspace of functions that are constant on each interval I_j. Then clearly for all $u \in L^2(\mathbb{R}_+)$, $j \geq 1$ and $t \in I_j$ we have $(P_h u)(t) = \bar{u}_{j,h}$.

Lemma 1. *Let $h > 0$, $\sigma = 2/h$, $T = Jh$ for some $J \in \mathbb{N}$, $u \in C_c(\mathbb{R}_+) \cap H^1(\mathbb{R}_+)$, and assume that* $\operatorname{supp}(u) := \{t \in \mathbb{R} : u(t) \neq 0\} \subset [0, T]$.

(i) *Then the sequence $\{c_j(h,s)\}_{j\geq 1}$ can be chosen so that $(L_\sigma - \mathcal{L})(I_{h,s}u)(s) = 0$ for all $s \in \overline{\mathbb{C}_+}$.*

(ii) *For any such choice of the sequence $\{c_j(h,s)\}_{j\geq 1}$, we have*

$$\begin{aligned}&|(L_\sigma u)(s) - (\mathcal{L}u)(s)| \\ &\leq \frac{hT^{1/2}|s|}{\pi}\left(\|I_{h,s}u - P_h u\|_{L^2(0,T)} + \frac{h}{\pi}|u|_{H^1(0,T)}\right)\end{aligned} \tag{3.2}$$

for all $s \in \overline{\mathbb{C}_+}$, where $|u|^2_{H^1(0,T)} = \int_0^T |u'(t)|^2\,dt$.

(iii) *The sequence* $\{c_j(h,s)\}_{j\geq 1}$ *in claim* (i) *can be chosen* optimally *so that*

$$\|I_{h,s}u - P_h u\|_{L^2(0,T)} \leq \frac{15}{218}\left(h^{-1/2}T^{-1/2} + \frac{|s|}{6e}\right)\|P_h u\|_{L^2(0,T)}$$

holds for a given $s \in i\mathbb{R}$, $T \geq 1$ *if* $9h \leq T^{2/3}e^{-\frac{4}{3}|s|T}$. *Furthermore, then*

$$\begin{aligned} &|(L_\sigma u)(s) - (\mathcal{L}u)(s)| \qquad (3.3)\\ &\leq \frac{3h^{1/2}|s|}{100}\|u\|_{L^2(0,T)} + \frac{2hT^{1/2}|s|^2}{1000}\|u\|_{L^2(0,T)} + \frac{h^2T^{1/2}|s|}{10}\|u\|_{H^1(0,T)}. \end{aligned}$$

Claim (iii) of this Lemma has an easy consequence that is easier to remember:

Corollary 1. *Under the assumption of Lemma 1, there exists a constant* $C < \infty$ *such that the estimate*

$$|(L_\sigma u)(i\omega) - (\mathcal{L}u)(i\omega)| < Ch^{1/2}(1+|\omega|^2)T^{1/2}\|u\|_{H^1(0,T)}$$

holds for all $T \geq 1$, $\omega \in \mathbb{R}$ *and* $0 < h < 1$ *satisfying* $9h \leq T^{2/3}e^{-\frac{4}{3}|\omega|T}$.

Proof of Lemma 1. Let us first make some general observations. By a simple argument, $\|P_h u\|^2_{L^2(\mathbb{R}_+)} = h\sum_{j\geq 1}\bar{u}^2_{j,h}$. Clearly for all $t \in I_j$

$$(I_{h,s}u - P_h u)(t) = \frac{c_j(h,s)}{h}(t - t_{j-1/2}).$$

Since for any $b > a$ we have

$$\frac{1}{(b-a)^2}\int_a^b \left(t - \frac{b+a}{2}\right)^2 dt = \frac{b-a}{12},$$

it follows that

$$\begin{aligned} \|I_{h,s}u - P_h u\|^2_{L^2(0,T)} &= \sum_{j=1}^{J}\frac{c_j(h,s)^2}{h^2}\int_{t_{j-1}}^{t_j}(t - t_{j-1/2})^2\,dt \qquad (3.4)\\ &= \frac{h}{12}\sum_{j=1}^{J}c_j(h,s)^2. \end{aligned}$$

In claim (i) we want to determine the sequence $\{c_j(h,s)\}_{j\geq 1}$ so as to satisfy $(L_\sigma - \mathcal{L})(I_{h,s}u)(s) = 0$ for given h and s. After some computations, we see that this is equivalent to requiring that $\{c_j(h,s)\}_{j\geq 1}$ satisfies

$$\sum_{j=1}^{J}\bar{u}_{j,h}I_j^{(0)}(h,s) + \sum_{j=1}^{J}c_j(h,s)J_j(h,s) = 0, \qquad (3.5)$$

where for $s \in \overline{\mathbb{C}_+} \setminus \{0\}$

$$I_j^{(0)}(h,s) := \int_{I_j} \left[\frac{1}{1+s/\sigma} \left(\frac{\sigma - s}{\sigma + s} \right)^j - e^{-st} \right] dt \tag{3.6}$$
$$= \frac{2}{\sigma + s} \left(\frac{\sigma - s}{\sigma + s} \right)^j + \frac{1}{s} \left[e^{-sjh} - e^{-s(j-1)h} \right]$$

and

$$J_j(h,s) := I_j^{(1)}(h,s) - (j - 1/2)h \cdot I_j^{(0)}(h,s) \tag{3.7}$$
$$= \frac{1}{s^2} \left[e^{-sjh} - e^{-s(j-1)h} \right] + \frac{h}{2s} \left[e^{-sjh} + e^{-s(j-1)h} \right]$$

together with

$$I_j^{(1)}(h,s) := \int_{I_j} \left[\frac{1}{1+s/\sigma} \left(\frac{\sigma - s}{\sigma + s} \right)^j - e^{-st} \right] t \, dt$$
$$= \frac{(2j-1)h}{\sigma + s} \left(\frac{\sigma - s}{\sigma + s} \right)^j + \left(\frac{jh}{s} + \frac{1}{s^2} \right) \left[e^{-sjh} - e^{-s(j-1)h} \right] + \frac{h}{s} e^{-s(j-1)h}.$$

It is clear that (3.5) has a huge number of solutions $\{c_j(h,s)\}_{j=1}^J$ for any fixed s and h, and most of the functions $(h,s) \mapsto c_j(h,s)$ need not even be continuous.

Claim (ii) is to be treated next. Recalling (2.6), (2.7) and (3.1)

$$\begin{aligned}
&(L_\sigma u)(s) - (\mathcal{L}u)(s) = \int_0^T (K_{s,\sigma}(t) - e^{-st}) u(t) \, dt \\
&= \int_0^T (K_{s,\sigma}(t) - e^{-st})(u(t) - (I_{h,s}u)(t)) \, dt \\
&= \sum_{j=1}^J \int_{t_{j-1}}^{t_j} (K_{s,\sigma}(t) - e^{-st})(u(t) - \bar{u}_{j,h}) \, dt \\
&- \sum_{j=1}^J \frac{c_j(h,s)}{h} \int_{t_{j-1}}^{t_j} (K_{s,\sigma}(t) - e^{-st})(t - t_{j-1/2}) \, dt =: \mathrm{I} - \mathrm{II}.
\end{aligned} \tag{3.8}$$

Let us first give an estimate to term II. By the Poincaré inequality (see, e.g., [8, Theorem 1.7]) we obtain for all $j = 1, \ldots, J$

$$\|(I - P_h)(K_{s,\sigma} - e^{-s(\cdot)})\|_{L^2(I_j)} \le \frac{h}{\pi} |K_{s,\sigma} - e^{-s(\cdot)}|_{H^1(I_j)} = \frac{h}{\pi} |e^{-s(\cdot)}|_{H^1(I_j)}$$

where the equality follows because the function $K_{s,\sigma}$ is constant on each interval I_j. By the mean value theorem we get for $s \in \mathbb{C}_+$ and $0 \le a < b < \infty$,

$$|e^{-s(\cdot)}|^2_{H^1(a,b)} = \int_a^b |\frac{d}{dt}e^{-st}|^2\,dt = \frac{|s|^2}{2\mathrm{Re}\,s}\left(e^{-2a\mathrm{Re}\,s} - e^{-2b\mathrm{Re}\,s}\right)$$
$$\le \frac{|s|^2}{2\mathrm{Re}\,s}\cdot 2\mathrm{Re}\,s e^{-2\xi\mathrm{Re}\,s}\,(b-a) \le (b-a)|s|^2 e^{-2a\mathrm{Re}\,s}.$$

Hence $|e^{-s(\cdot)}|_{H^1(I_j)} \le h^{1/2}|s|e^{-(j-1)h\mathrm{Re}\,s}$ and this estimate is seen to hold also for all $s \in \overline{\mathbb{C}_+}$. We now conclude that $|e^{-s(\cdot)}|_{H^1(0,T)} \le T^{1/2}|s|$ and

$$\|(I-P_h)(K_{s,\sigma} - e^{-s(\cdot)})\|_{L^2(I_j)} \le \frac{h^{3/2}|s|}{\pi} \tag{3.9}$$

for all $s \in \overline{\mathbb{C}_+}$. Using (3.9) we have

$$\begin{aligned}
\mathrm{II} &= \sum_{j=1}^J \int_{t_{j-1}}^{t_j} (K_{s,\sigma}(t) - e^{-st}) \cdot \frac{c_j(h,s)}{h}(t - t_{j-1/2})\,dt \qquad (3.10)\\
&= \sum_{j=1}^J \int_{t_{j-1}}^{t_j} \left((I-P_h)\left(K_{s,\sigma} - e^{-s(\cdot)}\right)\right)(t)\ \cdot \frac{c_j(h,s)}{h}(t - t_{j-1/2})\,dt\\
&\le \sum_{j=1}^J \frac{h^{3/2}|s|}{\pi}\cdot\left[\frac{c_j(h,s)^2}{h^2}\int_{t_{j-1}}^{t_j}(t-t_{j-1/2})^2\,dt\right]^{1/2}\\
&\le \left(\sum_{j=1}^J \frac{h^3|s|^2}{\pi^2}\right)^{1/2}\cdot\left(\sum_{j=1}^J \frac{c_j(h,s)^2}{h^2}\int_{t_{j-1}}^{t_j}(t-t_{j-1/2})^2\,dt\right)^{1/2}\\
&\le \frac{h^{3/2}|s|}{\pi}J^{1/2}\cdot\|I_{h,s}u - P_h u\|_{L^2(0,T)} = \frac{hT^{1/2}|s|}{\pi}\|I_{h,s}u - P_h u\|_{L^2(0,T)},
\end{aligned}$$

where the Schwarz inequality has been used twice, and the second to the last step is by (3.4).

It remains to estimate term I in (3.8). In this case, since P_h maps on piecewise constant functions and each $u(t) - \bar{u}_{j,h}$ has zero mean on subintervals

I_j, we obtain from (3.9) using the inequalities of Schwarz and Poincaré

$$\begin{aligned}
\mathrm{I} \leq & \sum_{j=1}^{J} \int_{t_{j-1}}^{t_j} \left((I - P_h) \left(K_{s,\sigma} - e^{-s(\cdot)} \right) \right)(t)(u(t) - \bar{u}_{j,h})\, dt \\
\leq & \sum_{j=1}^{J} \frac{h^{3/2}|s|}{\pi} \cdot \frac{h}{\pi} |u|_{H^1(I_j)} \leq \frac{h^{5/2}|s|}{\pi^2} \sum_{j=1}^{J} |u|_{H^1(I_j)} \qquad (3.11) \\
\leq & \frac{h^{5/2}|s|}{\pi^2} \left(\sum_{j=1}^{J} 1 \right)^{1/2} \left(\sum_{j=1}^{J} |u|^2_{H^1(I_j)} \right)^{1/2} = \frac{h^2 T^{1/2}|s|}{\pi^2} |u|_{H^1(0,T)}.
\end{aligned}$$

Estimate (3.2) follows from combining (3.10) and (3.11) with (3.8).

To prove claim (iii), we shall minimise $\frac{h}{12} \sum_{j \geq 1} c_j(h,s)^2$ under the constraint (3.5); see (3.4) for motivation. We form the Langrange function

$$\begin{aligned}
& L(c_1, \ldots, c_k \ldots, c_J, \lambda) \\
& = \frac{h}{12} \sum_{j=1}^{J} c_j^2 + \lambda \left(\sum_{j=1}^{J} \bar{u}_{j,h} I_j^{(0)}(h,s) + \sum_{j=1}^{J} c_j J_j(h,s) \right)
\end{aligned}$$

and compute its (unique) critical point giving the minimum. We obtain

$$\begin{cases} \frac{\partial L}{\partial c_k} = \frac{h}{6} c_k + \lambda J_k(h,s) = 0 \quad \text{for } 1 \leq k \leq J, \\ \sum_{j=1}^{J} \bar{u}_{j,h} I_j^{(0)}(h,s) + \sum_{j=1}^{J} c_j J_j(h,s) = 0. \end{cases}$$

Solving this gives the minimising sequence

$$c_k = c_k(h,s) = -\frac{6\lambda}{h} J_k(h,s) = -\frac{\sum_{j=1}^{J} \bar{u}_{j,h} I_j^{(0)}(h,s)}{\sum_{j=1}^{J} J_j(h,s)^2} J_k(h,s)$$

for all $1 \leq k \leq J$, and then for the minimum value

$$\begin{aligned}
& \frac{h}{12} \sum_{j=1}^{J} c_j(h,s)^2 = \frac{h}{12} \left(\frac{\sum_{j=1}^{J} \bar{u}_{j,h} I_j^{(0)}(h,s)}{\sum_{j=1}^{J} J_j(h,s)^2} \right)^2 \sum_{k=1}^{J} J_k(h,s)^2 \\
& = \frac{h}{12} \frac{\left(\sum_{j=1}^{J} \bar{u}_{j,h} I_j^{(0)}(h,s) \right)^2}{\sum_{j=1}^{J} J_j(h,s)^2}.
\end{aligned}$$

Hence, choosing the operator $I_{h,s}$ in (3.4) optimally gives

$$\|I_{h,s} u - P_h u\|_{L^2(0,T)} \leq \frac{\left(\sum_{j=1}^{J} I_j^{(0)}(h,s)^2 \right)^{1/2}}{\left(\sum_{j=1}^{J} J_j(h,s)^2 \right)^{1/2}} \frac{\|P_h u\|_{L^2([0],)}}{2\sqrt{3}}$$

since $\|P_h u\|_{L^2(0,T)} = \left(h\sum_{j=1}^{J} \bar{u}_{j,h}^2\right)^{1/2}$. We must now attack (3.6) and (3.7) to estimate the required two square sums, and the required long computations will be done in separate subsections 3.1 and 3.2 below. As a final result, we get by Propositions 5 and 6

$$\frac{\left(\sum_{j=1}^{J} I_j^{(0)}(h,s)^2\right)^{1/2}}{\left(\sum_{j=1}^{J} J_j(h,s)^2\right)^{1/2}} \le \frac{5}{218}\left(3h^{-1/2}T^{-1/2} + h^{1/2}|s|^2 T^{1/2}\right)$$

assuming that $9h \le T^{2/3}e^{-\frac{4}{3}|s|T}$. But then

$$h^{1/2}|s|^2T^{1/2} \le \frac{|s|}{3}\cdot|s|T^{5/6}e^{-\frac{2}{3}|s|T} \le \frac{|s|}{3}\cdot|s|Te^{-\frac{2}{3}|s|T} \le \frac{|s|}{2e}$$

since $\max_{r\ge 0} re^{-\frac{2}{3}r} = 3/(2e)$. Noting that the norm of the orthogonal projection P_h is 1, the proof of Lemma 1 is now complete. □

3.1 Estimation of (3.7)

In this subsection, we shall estimate the square sum of

$$J_j(h,s) = \frac{1}{s^2}\left[e^{-sjh} - e^{-s(j-1)h}\right] + \frac{h}{2s}\left[e^{-sjh} + e^{-s(j-1)h}\right] \tag{3.12}$$

from below and above. For the first term on the left of (3.12) we obtain

$$\begin{aligned}
&\frac{1}{s^2}\left[e^{-sjh} - e^{-s(j-1)h}\right] = \frac{1}{s^2}\left[\sum_{k\ge 0}\frac{(-sjh)^k}{k!} - \sum_{k\ge 0}\frac{(-s(j-1)h)^k}{k!}\right]\\
&= \frac{1}{s^2}\left[-sh + \sum_{k\ge 2}\frac{(-sh)^k(j^k-(j-1)^k)}{k!}\right]\\
&= -\frac{h}{s} + \sum_{k\ge 2}\frac{(j^k-(j-1)^k)}{k!}(-s)^{k-2}h^k.
\end{aligned}$$

For the latter term in (3.12) we get

$$\begin{aligned}
&\frac{h}{2s}\left[e^{-sjh} + e^{-s(j-1)h}\right] = \frac{h}{s}\sum_{k\ge 0}\frac{(-s)^k(j^k+(j-1)^k)}{2k!}h^k\\
&= \frac{h}{s} - \sum_{k\ge 2}\frac{(j^{k-1}+(j-1)^{k-1})}{2(k-1)!}(-s)^{k-2}h^k.
\end{aligned}$$

Hence, for all $s \in \overline{\mathbb{C}_+} \setminus \{0\}$

$$J_j(h,s) = \sum_{k\geq 2} \frac{d_k(j)}{2k!}(-s)^{k-2}h^k,$$

where the coefficient polynomials satisfy (by the binomial theorem)

$$\begin{aligned} d_k(j) &= 2\left(j^k - (j-1)^k\right) - k\left(j^{k-1} + (j-1)^{k-1}\right) \\ &= \sum_{m=0}^{k-3} \binom{k}{m} (k-m-2)(-1)^{k-m} j^m \quad \text{for} \quad k \geq 3 \end{aligned}$$

and $d_2(j) = 0$. Hence $d_k(j)$ is a polynomial of degree $k-3$ in variable j. Finally, we get the expression

$$J_j(h,s) = \sum_{k\geq 3} \sum_{m=0}^{k-3} \frac{k-m-2}{2m!(k-m)!}(-j)^m s^{k-2}h^k.$$

Let us compute an upper estimate for $\|\{J_j(h,s)\}_j\|_{\ell^2} := \left(\sum_{j=1}^J J_j(h,s)^2\right)^{1/2}$. By the triangle inequality

$$\begin{aligned} &\|\{J_j(h,s)\}_j\|_{\ell^2} \\ &\leq |s^{-2}| \cdot \sum_{k\geq 3} \sum_{m=0}^{k-3} \frac{k-m-2}{2m!(k-m)!}|sh|^k \left(\sum_{j=1}^J j^{2m}\right)^{1/2} \\ &\leq |s^{-2}| \cdot \sum_{k\geq 3} \sum_{m=0}^{k-3} \frac{k-m-2}{2m!(k-m)!}|sh|^k \cdot \frac{J^{m+1/2}}{\sqrt{2m+1}} \\ &\leq \frac{1}{2}|s|T^{1/2}h^{5/2} \cdot \sum_{k\geq 3} \sum_{m=0}^{k-3} \frac{k-m-2}{2\sqrt{2m+1}\, m!(k-m)!}|s|^{k-3}T^m h^{k-m-3}. \end{aligned}$$

Noting that for $k-3 \geq m \geq 0$ we have $\frac{k-m-2}{\sqrt{2m+1}\, m!(k-m)!} \leq \frac{1}{m!(k-m-3)!}$ and $|s|^{k-3}T^m h^{k-m-3} = |sh|^{k-3} \cdot (T/h)^m$, we may estimate the sum term above

$$\begin{aligned} &\sum_{k\geq 3} \sum_{m=0}^{k-3} \frac{k-m-2}{2\sqrt{2m+1}\, m!(k-m)!}|s|^{k-3}T^m h^{k-m-3} \\ &\leq \sum_{k\geq 3} \left(\frac{|sh|^{k-3}}{(k-3)!} \sum_{m=0}^{k-3} \binom{k-3}{m} \left(\frac{T}{h}\right)^m \right) \\ &\leq \sum_{k\geq 3} \frac{|sh|^{k-3}}{(k-3)!} \left(1 + \frac{T}{h}\right)^{k-3} = e^{|s|(h+T)}. \end{aligned}$$

We now conclude for all $h, T > 0$ and $s \in \overline{\mathbb{C}_+} \setminus \{0\}$ that

$$\|\{J_j(h,s)\}_{j=1}^J\|_{\ell^2} \leq \frac{1}{2}|s|T^{1/2}h^{5/2}e^{|s|(h+T)}. \tag{3.13}$$

In addition to estimate (3.13) a lower bound can also be obtained. Decompose

$$\begin{aligned} J_j(h,s) &= \sum_{k=3}^{\infty}\sum_{m=0}^{k-3}\frac{k-m-2}{2m!(k-m)!}(-j)^m s^{k-2}h^k \\ &= \sum_{k=3}^{\infty}\left(\frac{1}{2(k-3)!3!}(-j)^{k-3}s^{k-2}h^k + \sum_{m=0}^{k-4}\frac{k-m-2}{2m!(k-m)!}(-j)^m s^{k-2}h^k\right) \\ &= \sum_{k=3}^{\infty}\frac{1}{2(k-3)!3!}(-j)^{k-3}s^{k-2}h^k + \sum_{k=4}^{\infty}\sum_{m=0}^{k-4}\frac{k-m-2}{2m!(k-m)!}(-j)^m s^{k-2}h^k \end{aligned}$$

so that by the triangle inequality

$$\begin{aligned} \left\|\{J_j(h,s)\}_{j=1}^J\right\|_{\ell^2} \geq{}& \left\|\left\{\sum_{k=3}^{\infty}\frac{1}{2(k-3)!3!}(-j)^{k-3}s^{k-2}h^k\right\}_{j=1}^J\right\|_{\ell^2} \\ &- \left\|\left\{\sum_{k=4}^{\infty}\sum_{m=0}^{k-4}\frac{k-m-2}{2m!(k-m)!}(-j)^m s^{k-2}h^k\right\}_{j=1}^J\right\|_{\ell^2}. \end{aligned} \tag{3.14}$$

For the first term in the right hand side of (3.14) we have

$$\begin{aligned} &\left\|\left\{\sum_{k=3}^{\infty}\frac{1}{2(k-3)!3!}(-j)^{k-3}s^{k-2}h^k\right\}_{j=1}^J\right\|_{\ell^2} \\ &= \left\|\left\{\frac{1}{12}sh^3\sum_{k=3}^{\infty}\frac{1}{(k-3)!}(-j)^{k-3}s^{k-3}h^{k-3}\right\}_{j=1}^J\right\|_{\ell^2} \\ &= \frac{1}{12}|s|h^3\cdot\left\|\left\{e^{-jsh}\right\}_{j=1}^J\right\|_{\ell^2}, \end{aligned} \tag{3.15}$$

where

$$\begin{aligned} \left\|\left\{e^{-jsh}\right\}_{j=1}^J\right\|_{\ell^2}^2 &= \sum_{j=1}^J |e^{-jsh}|^2 \\ &= \begin{cases} J = h^{-1}T, & \text{when } \operatorname{Re} s = 0 \\ e^{-2h\operatorname{Re} s}\frac{1-e^{-2(J+1)h\operatorname{Re} s}}{1-e^{-2h\operatorname{Re} s}}, & \text{when } \operatorname{Re} s > 0. \end{cases} \end{aligned} \tag{3.16}$$

For the latter term in (3.14) we have a similar upper estimate to (3.13). Indeed,

$$\begin{aligned}
&\Big\| \Big\{ \sum_{k=4}^{\infty} \sum_{m=0}^{k-4} \frac{k-m-2}{2m!(k-m)!} (-j)^m s^{k-2} h^k \Big\}_{j=1}^{J} \Big\|_{\ell^2} \\
\leq & \sum_{k=4}^{\infty} \sum_{m=0}^{k-4} \frac{k-m-2}{2m!(k-m)!} |s|^{k-2} h^k \frac{J^{m+1/2}}{\sqrt{2m+1}} \\
= & \sum_{k=4}^{\infty} \sum_{m=0}^{k-4} \frac{k-m-2}{2m!(k-m)!} |s|^{k-2} h^k h^{-m-1/2} T^{m+1/2} \\
= & |s|^2 h^{7/2} \sum_{k=4}^{\infty} \sum_{m=0}^{k-4} \frac{k-m-2}{2m!(k-m)!} |s|^{k-4} h^{k-m-4} T^m \\
\leq & |s| h^{7/2} e^{|s|(h+T)}.
\end{aligned} \tag{3.17}$$

As a conclusion we can now state the following proposition:

Proposition 5. *Let $J_j(h,s)$ be defined through (3.12). Then for any $s \in i\mathbb{R}$, $T, h > 0$ satisfying $T = Jh$, $J \in \mathbb{N}$ and $9h \leq T^{2/3} e^{-\frac{4}{3}|s|T}$ we have*

$$\|\{J_j(h,s)\}_{j=1}^{J}\|_{\ell^2} \geq \frac{5}{109} T h^2 |s|. \tag{3.18}$$

Proof. It is clear that (3.18) is satisfied for $s = 0$. For $s \in i\mathbb{R} \setminus \{0\}$ it follows from (3.14) and (3.15) – (3.17) that for all $s \in i\mathbb{R} \setminus \{0\}$, $h, T > 0$ satisfying $T = Jh$ for $J \in \mathbb{N}$ that the estimate

$$\big\|\{J_j(h,s)\}_{j=1}^{J}\big\|_{\ell^2} \geq \left(\frac{T}{12} - h^{3/2} e^{|s|(h+T)} \right) h^2 |s|$$

holds. Since always $h \leq T$, we have $h^{3/2} e^{|s|(h+T)} \leq h^{3/2} e^{2|s|T} \leq \frac{T}{27}$ provided that $h \leq \frac{T^{2/3}}{9} e^{-\frac{4}{3}|s|T}$. The claim follows from this. □

3.2 Estimation of (3.6)

In this subsection, we compute an upper estimate for $\left\|\left\{I_j^{(0)}(h,s)\right\}_{j=1}^{J}\right\|_{\ell^2} := \left(\sum_{j=1}^{J} I_j^{(0)}(h,s)^2\right)^{1/2}$. Writing $\tau = sh$ and $\sigma = 2/h$, we get for $s \in \overline{\mathbb{C}_+}$

$$\begin{aligned} I_j^{(0)}(h,s) =& \frac{2}{\sigma+s}\left(\frac{\sigma-s}{\sigma+s}\right)^j + \frac{1}{s}\left(e^{-sjh} - e^{-s(j-1)h}\right) \\ =& \frac{2}{\sigma+s}\left(\left(\frac{\sigma-s}{\sigma+s}\right)^j - e^{-sjh}\right) + \left(\frac{2}{\sigma+s} - \frac{1}{s}(e^{sh}-1)\right)e^{-sjh} \\ =& \frac{2h}{2+\tau}\left(\left(\frac{2-\tau}{2+\tau}\right)^j - e^{-\tau j}\right) + \left(\frac{2h}{2+\tau} - \frac{h}{\tau}(e^{\tau}-1)\right)e^{-\tau j}. \end{aligned}$$

Let $\Omega \subset \overline{\mathbb{C}_+}$ be any set. Then for any $\tau \in \Omega$ we have

$$\begin{aligned} |I_j^{(0)}(h,s)| \leq& \left|\frac{2h}{2+\tau}\right|\left|\left(\frac{2-\tau}{2+\tau}\right)^j - e^{-\tau j}\right| + \left|\frac{2h}{2+\tau} - \frac{h}{\tau}(e^{\tau}-1)\right|\left|e^{-\tau j}\right| \\ \leq& \left|\frac{2h}{2+\tau}\right|\left|\left(\frac{2-\tau}{2+\tau}\right) - e^{-\tau}\right|\left|\sum_{k=1}^{j-1}\left(\frac{2-\tau}{2+\tau}\right)^k e^{-\tau(j-k-1)}\right| \\ &+ \left|\frac{2h}{2+\tau} - \frac{h}{\tau}(e^{\tau}-1)\right| \\ \leq& h|\tau|\left(C_\Omega\left|\frac{2j\tau^2}{2+\tau}\right| + C'_\Omega\right), \end{aligned}$$

where the constants are given by

$$C_\Omega = \sup_{\tau\in\Omega}\left|\frac{1}{\tau^3}\left(\frac{2-\tau}{2+\tau} - e^{-\tau}\right)\right| \text{ and } C'_\Omega = \sup_{\tau\in\Omega}\left|\frac{1}{\tau}\left(\frac{2}{2+\tau} - \frac{1}{\tau}(e^{\tau}-1)\right)\right|.$$

This implies for all $h \geq 0$ and $\tau = sh \in \Omega$

$$\begin{aligned} \left\|\left\{I_j^{(0)}(h,s)\right\}_{j=1}^{J}\right\|_{\ell^2} \leq& C_\Omega \frac{2h|\tau|^3}{|2+h|}\left(\sum_{j=1}^{J} j^2\right)^{1/2} + C'_\Omega h|\tau|\left(\sum_{j=1}^{J} 1\right)^{1/2} \\ \leq& C_\Omega h^4|s|^3\left(\frac{1}{3}J^3 + \frac{1}{2}J^2 + \frac{1}{6}J\right)^{1/2} + C'_\Omega h^2|s|J^{1/2} \quad (3.19) \\ \leq& C_\Omega h^{5/2}|s|^3 T^{3/2} + C'_\Omega h^{3/2}|s|T^{1/2} \end{aligned}$$

by the facts that $T = Jh$ and $J \geq 1$. We now have to choose the set Ω in a clever way, so that the resulting estimate is properly "fine tuned" according to Proposition 5.

Proposition 6. *Let $I_j^{(0)}(h,s)$ be defined through (3.6). Then for any $s \in i\mathbb{R}$, $T \geq 1$,$h > 0$ satisfying $T = Jh$, $J \in \mathbb{N}$ and $9h \leq T^{2/3}e^{-\frac{4}{3}|s|T}$ we have*

$$\left\| \left\{ I_j^{(0)}(h,s) \right\}_{j=1}^{J} \right\|_{\ell^2} \leq \frac{1}{2}h^{5/2}|s|^3T^{3/2} + \frac{3}{2}h^{3/2}|s|T^{1/2}. \tag{3.20}$$

Proof. Since we assume (motivated by Proposition 5) that $9h \leq T^{2/3}e^{-\frac{4}{3}|s|T}$, we have

$$|\tau| = |s|h \leq \frac{|s|T^{2/3}}{9}e^{-\frac{4}{3}|s|T} \leq \frac{|s|T}{9}e^{-\frac{4}{3}|s|T} \leq \frac{1}{12e}$$

since $\max_{r\geq 0} re^{-\frac{4}{3}r} = 3/(4e)$. Hence, we must estimate the constants C_Ω and C'_Ω for the set $\Omega := [-i/(12e), i/(12e)]$. By computing the Taylor series, we see that

$$C_\Omega \leq \sum_{j\geq 0} \left| \frac{1}{2^{j+2}} - \frac{1}{(j+3)!} \right| \cdot \left(\frac{1}{12e} \right)^j < \sum_{j\geq 0} \frac{1}{2^{j-1}} \cdot \left(\frac{1}{12e} \right)^j < \frac{1}{2}$$

and similarly

$$C'_\Omega \leq \sum_{j\geq 0} \left| \left(-\frac{1}{2} \right)^{j+1} - \frac{1}{(j+2)!} \right| \cdot \left(\frac{1}{12e} \right)^j < \sum_{j\geq 0} \frac{1}{2^j} \cdot \left(\frac{1}{12e} \right)^j < \frac{3}{2}.$$

But now (3.19) implies (3.20). □

4 Weak and strong convergence

Our main results are given in this section. We first show that Lemma 1 implies that $L_\sigma \to \mathcal{L}$ in weak operator topology. Using this, it is then shown in Theorem 1 that the convergence is actually strong. The input/output approximation of linear dynamical systems is treated in Theorem 2.

It follows from Lemma 1 that $(L_\sigma u)(i\omega) \to (\mathcal{L}u)(i\omega)$ uniformly in the compact subsets $i\omega \in K \subset i\mathbb{R}$ for any $u \in C_c(\mathbb{R}_+) \cap H^1(\mathbb{R}_+)$. Hence, for finite linear combinations s of characteristic functions χ_K of compact intervals $K \subset i\mathbb{R}$ (also called simple functions) we have $\langle s, L_\sigma u\rangle_{L^2(i\mathbb{R})} \to \langle s, \mathcal{L}u\rangle_{L^2(i\mathbb{R})}$. Since $\|L_\sigma\|_{\mathcal{L}(L^2(\mathbb{R}_+);H^2(\mathbb{C}_+))} \leq 1$ and simple functions are dense in $L^2(i\mathbb{R})$, it follows that

$$\langle v, L_\sigma u\rangle_{K^2(i\mathbb{R})} \to \langle v, \mathcal{L}u\rangle_{H^2(i\mathbb{R})} \text{ as } \sigma \to \infty \tag{4.1}$$

for all $u \in C_c(\mathbb{R}) \cap H^1(\mathbb{R}_+)$ and $v \in L^2(i\mathbb{R}_+)$. Another density argument implies finally that (4.1) holds even for all $u \in L^2(\mathbb{R}_+)$ and $v \in L^2(i\mathbb{R}_+)$. To continue the argument, we recall a result from elementary functional analysis:

Proposition 7. *Let H be a Hilbert space, and assume that $u_j \to u$ weakly in H. If $\|u_j\|_H \to \|u\|_H$, then $u_j \to u$ in the norm of H.*

Proof. $\langle u_j - u, u_j - u\rangle_H = \langle u_j, u_j\rangle_H - \langle u, u\rangle_H - \langle u, u_j - u\rangle_H - \langle u_j - u, u\rangle_H = \|u_j\|_H^2 - \|u\|_H^2 - 2\mathrm{Re}\,\langle u, u_j - u\rangle_H$. □

Theorem 1. *We have $\|L_\sigma u - \mathcal{L}u\|_{H^2(\mathbb{C}_+)} \to 0$ for any $u \in L^2(\mathbb{R}_+)$. Moreover, $\|L_\sigma^* v - \mathcal{L}^* v\|_{L^2(\mathbb{R}_+)} \to 0$ for any $v \in H^2(\mathbb{C}_+)$.*

Proof. Adjoining (4.1) shows that $L_\sigma^* v \to \mathcal{L}^* v$ weakly. Since L_σ is a coisometry by Proposition 3 and (2.4), we have

$$\|L_\sigma^* v\|_{L^2(\mathbb{R}_+)}^2 = \langle L_\sigma L_\sigma^* v, v\rangle_{H^2(\mathbb{C}_+)}^2 = \|v\|_{H^2(\mathbb{C}_+)}^2.$$

Now Proposition 7 implies the latter part of this Theorem.

To show the first part, we have to work a bit harder to verify that $\|L_\sigma u\|_{L^2(i\mathbb{R})} \to \|u\|_{L^2(\mathbb{R}_+)} = \|\mathcal{L}u\|_{L^2(i\mathbb{R})}$. Suppose that $h = 2/\sigma > 0$ and $u \in L^2(\mathbb{R}_+)$ is such that $u(t) = \overline{u}_{j,h} := \int_{((j-1)h, jh]} u(t)\,dt$ for all $t \in I_j := ((j-1)h, jh]$ — in other words, this is simply $u = P_h u$. For such u

$$\|u\|_{L^2(\mathbb{R}_+)}^2 = \sum_{j\geq 1} \int_{I_j} |u(t)|^2\,dt = h\|\{\overline{u}_{j,h}\}_{j\geq 0}\|_{\ell^2}^2.$$

By the definition of the discretising operator T_σ, we have

$$\|T_\sigma u\|_{H^2(\mathbb{D})}^2 = \sum_{j\geq 1} \left(\frac{1}{\sqrt{h}} \int_{I_j} |u(t)|^2\,dt\right)^2 = h \sum_{j\geq 1} |\overline{u}_{j,h}|^2 = \|u\|_{L^2(\mathbb{R}_+)}^2.$$

Hence, we have $\|T_\sigma P_h u\|_{H^2(\mathbb{D})} = \|P_h u\|_{L^2(\mathbb{R}_+)}$ for all $u \in L^2(\mathbb{R}_+)$ where $\sigma = 2/h$. Also note that $T_\sigma u = T_\sigma P_h u$ for all $u \in L^2(\mathbb{R}_+)$ provided that $\sigma = 2/h$. We now have for any $u \in L^2(\mathbb{R}_+)$

$$\begin{aligned}
&\left|\|T_\sigma u\|_{H^2(\mathbb{D})} - \|u\|_{L^2(\mathbb{R}_+)}\right| \\
&\leq \left|\|T_\sigma u\|_{H^2(\mathbb{D})} - \|T_\sigma P_h u\|_{H^2(\mathbb{D})}\right| + \left|\|T_\sigma P_h u\|_{H^2(\mathbb{D})} - \|P_h u\|_{L^2(\mathbb{R}_+)}\right| \\
&+ \left|\|P_h u\|_{L^2(\mathbb{R}_+)} - \|u\|_{L^2(\mathbb{R}_+)}\right| = \left|\|P_h u\|_{L^2(\mathbb{R}_+)} - \|u\|_{L^2(\mathbb{R}_+)}\right|,
\end{aligned}$$

where again $\sigma = 2/h$. Since the projections $P_h \to I$ strongly in $L^2(\mathbb{R}_+)$ as $h \to 0$, we conclude that $\|T_\sigma u\|_{H^2(\mathbb{D})} \to \|u\|_{L^2(\mathbb{R}_+)}$ and hence $\|L_\sigma u\|_{H^2(\mathbb{C}_+)} \to \|u\|_{L^2(\mathbb{R}_+)}$ as $\sigma \to \infty$, see Proposition 3. The first claim of this theorem follows from this, Proposition 7 and equation (4.1). □

Using Theorem 1 we can finally show that the output of integration scheme (1.4) converges to the output of continuous time dynamics (1.1) for input/output stable systems S.

Theorem 2. *For any $u \in L^2(\mathbb{R}_+)$ and $\widehat{\mathcal{G}} \in H^\infty(\mathbb{C}_+)$, we have*

$$\|T_\sigma^* \widehat{\mathcal{D}}_\sigma T_\sigma u - \mathcal{L}^* \widehat{\mathcal{G}} \mathcal{L} u\|_{L^2(\mathbb{R}_+)} \to 0 \tag{4.2}$$

as $\sigma \to \infty$.

Proof. As noted just before Proposition 4, we have $T_\sigma^* \widehat{\mathcal{D}}_\sigma T_\sigma = L_\sigma^* \widehat{\mathcal{G}} L_\sigma$. Then we get for all $\sigma > 0$

$$\begin{aligned}\|L_\sigma^* \widehat{\mathcal{G}} L_\sigma u - \mathcal{L}^* \widehat{\mathcal{G}} \mathcal{L} u\|_{L^2(\mathbb{R}_+)} &\le \|(L_\sigma^* - \mathcal{L}^*)\, \widehat{\mathcal{G}}\, (L_\sigma u - \mathcal{L} u)\|_{L^2(\mathbb{R}_+)} \\ &+ \|(L_\sigma^* - \mathcal{L}^*)\, \widehat{\mathcal{G}} \mathcal{L} u\|_{L^2(\mathbb{R}_+)} + \|\mathcal{L}^* \widehat{\mathcal{G}}\, (L_\sigma u - \mathcal{L} u)\|_{L^2(\mathbb{R}_+)}.\end{aligned}$$

Now (4.2) follows by Theorem 1. □

5 On the optimality of Theorem 2

We complete this paper by showing that Theorem 2 is optimal in the sense that it cannot be improved to have a speed estimate for convergence as in Lemma 1. To this end, we consider estimate (2.5) in the special case when $\widehat{\mathcal{G}}(s) = I$ for all $s \in \mathbb{C}_+$.

In this special case it follows from the very definitions that $L_\sigma^* \widehat{\mathcal{G}} L_\sigma = T_\sigma^* T_\sigma = P_{2/\sigma}$ where the orthogonal projection P_h is defined as in Section 3. Since $\mathcal{L}^* \mathcal{L} = \mathcal{I}$ on all of $L^2(\mathbb{R}_+)$, we should give an estimate to

$$\|u - P_h u\|_{L^2(0,T)} \quad \text{for a family of functions} \quad u \in L^2(\mathbb{R}_+).$$

It is, of course, true that $P_h u \to u$ as $h \to 0$ for all $u \in L^2(\mathbb{R}_+)$. However, there cannot be a uniform speed estimate of type

$$\|u - P_h u\|_{L^2(0,T)} \le C_u h^\alpha, \tag{5.1}$$

where $C_u < \infty$ for *all* $u \in L^2(0,T)$. If it were so, then for any $0 < \beta < \alpha$ we would have $\|h^{-\beta}(I - P_h)u\|_{L^2(0,T)} \le C_u h^{\alpha-\beta} \to 0$ as $h \to 0$, for all $u \in L^2(0,T)$. By the uniform boundedness principle,

$$\sup_{h>0} \|h^{-\beta}(I - P_h)\|_{L^2(0,T)} =: M < \infty$$

and hence $\|(I - P_h)\|_{\mathcal{L}(L^2(0,T))} \le M h^\beta$ for all $h > 0$.

Making now h small enough, we see that then the norm of the orthogonal projection $(I - P_h)|L^2(0,T)$ is strictly less than 1; this implies that $I|L^2(0,T) = P_h|L^2(0,T)$. But $P_h|L^2(0,T)$ is a finite rank operator, and the

uniform speed estimate (5.1) cannot hold by contradiction. The same conclusion holds, if h^α in (5.1) is replaced by *any* increasing continuous function $\phi(h)$ satisfying $\phi(0) = 0$.

It should be noted that a speed estimate of type (5.1) can be obtained for functions $u \in L^2(\mathbb{R}_+)$ that have some "smoothness". See [4] for a further discussion on what is obtainable and what is not.

6 Conclusions and remarks

We have shown in Section 1 that the Cayley transform (in the context of linear system theory) is equivalent to the classical Tustin discretisation (1.2) even for infinite-dimensional linear systems $S = \left[\begin{smallmatrix} A\&B \\ C\&D \end{smallmatrix}\right]$. The convergence of this discretisation is studied in the scalar-valued input/output setting, using the operators L_σ as introduced before Proposition 4.

It is shown in Theorem 1 (see also Corollary 1) that for a wide class of functions u, the function $L_\sigma u$ provides a pointwise approximation to the usual Laplace transform. Even a convergence speed estimate is given as a function of the sampling parameter $h = 2/\sigma$. This result is extended to the input/output mapping of the linear system S; see Theorem 2.

Unfortunately, Theorem 2 cannot be improved with a speed estimate, as discussed in Section 5. This is understandable since for any $\sigma > 0$, the sampling operator T_σ cannot detect above a certain cutoff frequency. However, there are always high frequency signals u carrying substantial energy that the discretised input/output mapping $T_\sigma^* \widehat{\mathcal{D}}_\sigma T_\sigma$ of S cannot capture at all.

It is possible to make some variants of Theorem 2 to operator-valued transfer functions $\widehat{\mathcal{G}}(\cdot)$ but we do not discuss them here. Likewise, the approximation of the true state trajectory $x(\cdot)$ in (1.11) by the discrete trajectories $\{x_j^{(h)}\}_{j\geq 0}$ solving (1.4) remains a subject of further study.

Acknowledgment

We would like to thank several anonymous reviewers for valuable comments.

References

[1] D. Arov and I. Gavrilyuk. A method for solving initial value problems for linear differential equations in Hilbert space based on the Cayley transform. *Numerical Functional Analysis and Optimization*, 14(5&6):459–473, 1993.

[2] D. Z. Arov and M. A. Nudelman. Passive linear stationary dynamical scattering systems with continuous time. *Integral Equations Operator Theory*, 24:1–45, 1996.

[3] J. Ball and O. J. Staffans. Conservative state-space realizations of dissipative system behaviors. *Integral Equations Operator Theory*, 54:151–213, 2006.

[4] D. Braess. *Finite elements. Theory, fast solvers and applications in solid mechanics*. Cambridge University Press, Cambridge, 2001.

[5] M. S. Brodskiĭ. On operator colligations and their characteristic functions. *Soviet Mat. Dokl.*, 12:696–700, 1971.

[6] M. S. Brodskiĭ. *Triangular and Jordan representations of linear operators*, volume 32. American Mathematical Society, Providence, Rhode Island, 1971.

[7] M. S. Brodskiĭ. Unitary operator colligations and their characteristic functions. *Russian Math. Surveys*, 33(4):159–191, 1978.

[8] B. Dacorognan. *Direct Methods in the Calculus of Variations*. Springer Verlag, Berlin, 1989.

[9] I. P. Gavrilyuk. An algorithmic representation of fractional powers of positive operators. *Numerical Functional Analysis and Optimization*, 17(3-4):293–305, 1996.

[10] I. P. Gavrilyuk. A class of fully discrete approximations for the first order differential equations in Banach spaces with uniform estimates on the whole of $\mathbb{R}^+$. *Numerical Functional Analysis and Optimization*, 20(7&8):675–693, 1999.

[11] I. P. Gavrilyuk. Strongly P-positive operators and explicit representations of the solutions of initial value problems for second-order differential equations in Banach space. *Journal of Mathematical Analysis and Applications*, 236(2):327–349, 1999.

[12] I. P. Gavrilyuk and V. L. Makarov. The Cayley transform and the solution of an initial value problem for a first order differential equation with an unbounded operator coefficient in Hilbert space. *Numerical Functional Analysis and Optimization*, 15(5&6):583–598, 1994.

[13] I. P. Gavrilyuk and V. L. Makarov. Representation and approximation for the solution of second order differential equations with unbounded operator coefficients in Hilbert space. *Zeitschrift für Angewandte Mathematik und Mechanik*, 76(2):527–528, 1996.

[14] I. P. Gavrilyuk and V. L. Makarov. Representation and approximation of the solution of an initial value problem for a first order differential equation in Banach spaces. *Zeitschrift für Analysis und ihre Anwendungen. Journal for Analysis and its Applications*, 15(2):495–527, 1996.

[15] I. P. Gavrilyuk and V. L. Makarov. Exact and approximate solutions of some operator equations based on the Cayley transform. *Linear Algebra and its Applications*, 282(1-3):97–121, 1998.

[16] I. P. Gavrilyuk and V. L. Makarov. Explicit and approximate solutions of second order elliptic differential equations in Hilbert and Banach spaces. *Numerical Functional Analysis and Optimization*, 20(7&8):695–715, 1999.

[17] I. P. Gavrilyuk and V. L. Makarov. Explicit and approximate solutions of second-order evolution differential equations in Hilbert space. *Numerical Methods for Partial Differential Equations*, 15(1):111–131, 1999.

[18] I. P. Gavrilyuk and V. L. Makarov. Algorithms without accuracy saturation for evolution equations in Hilbert and Banach spaces. *Mathematics of Computation*, 74(250):555–583, 2004.

[19] V. I. Gorbachuk and M. L. Gorbachuk. *Boundary value problems for operator differential equations*, volume 48 of *Mathematics and its Applications (Soviet Series)*. Kluwer Academic Publishers Group, Dordrecht, 1991.

[20] E. Hairer, C. Lubich, and G. Wanner. *Geometric Numerical Integration: Structure-Preserving Algorithms for Ordinary Differential Equations*. Springer Verlag, Berlin, 2002.

[21] V. Havu and J. Malinen. Laplace and Cayley transforms – an approximation point of view. In *Proceedings of the Joint 44th IEEE Conference on Decision and Control and European Control Conference (CDC-ECC'05)*, Seville, 2005.

[22] M. S. Livšic and A. A. Yantsevich. *Operator colligations in Hilbert space*. John Wiley & sons, Inc., New York, 1977.

[23] M.S. Livšic. *Operators, vibrations, waves. Open systems.* Nauka, Moscow, 1966.

[24] J. Malinen. Conservativity and time-flow invertiblity of boundary control systems. In *Proceedings of the Joint 44th IEEE Conference on Decision and Control and European Control Conference (CDC-ECC'05)*, Seville, 2005.

[25] J. Malinen, O. Staffans, and G. Weiss. When is a linear system conservative? *Quarterly of Applied Mathematics*, 64:61–91, 2006.

[26] J. Malinen and O. J. Staffans. Conservativity of boundary control systems. *Journal of Differential Equations*, 231(1):290–312, 2006.

[27] J. Malinen and O. J. Staffans. Impedance passive and conservative boundary control systems. *Complex Analysis and Operator Theory*, 2(1), 2007.

[28] R. Ober and S. Montgomery-Smith. Bilinear transformation of infinite-dimensional state-space systems and balanced realizations of nonrational transfer functions. *SIAM Journal of Control and Optimization*, 28(2):438–465, 1990.

[29] J. D. Powell G. F. Franklin and M. L. Workman. *Digital Control of Dynamic Systems.* Addison-Wesley Publishing Company, Reading, Massachusetts, 1990.

[30] Yu. L. Smuljan. Invariant subspaces of semigroups and Lax–Phillips scheme. *Dep. in VINITI, No. 8009-B86, Odessa (A private translation by Daniela Toshkova, 2001)*, 1986.

[31] O. J. Staffans. J-energy preserving well-posed linear systems. *International Journal of Applied Mathematics and Computer Science*, 11:1361–1378, 2001.

[32] O. J. Staffans. Passive and conservative continuous-time impedance and scattering systems. Part I: Well-posed systems. *Mathematics of Control, Signals, and Systems*, 15:291–315, 2002.

[33] O. J. Staffans. *Well-Posed Linear Systems.* Cambridge University Press, Cambridge, 2004.

[34] O. J. Staffans and G. Weiss. Transfer functions of regular linear systems, Part II: The system operator and the Lax – Phillips semigroup.

Transactions of the American Mathematical Society, 354(8):3229–3262, 2002.

[35] O. J. Staffans and G. Weiss. Transfer functions of regular linear systems, Part III: Inversions and duality. *Integral Equations Operator Theory*, 49(4):517–558, 2004.

[36] B. Sz.-Nagy and C. Foias. *Harmonic Analysis of Operators on Hilbert space*. North-Holland Publishing Company, Amsterdam, 1970.

[37] M. Tucsnak and G. Weiss. How to get a conservative well-posed linear system out of thin air. II. Controllability and stability. *SIAM Journal on Control and Optimization*, 42(3):907–935, 2003.

[38] G. Weiss. Transfer functions of regular linear systems, Part I: Characterizations of regularity. *Transactions of American Mathematical Society*, 342(2):827–854, 1994.

[39] G. Weiss and M. Tucsnak. How to get a conservative well-posed linear system out of thin air. I. Well-posedness and energy balance. *ESAIM. Control, Optimisation and Calculus of Variations*, 9:247–274, 2003.