

MODEL REDUCTION BY BALANCED TRUNCATION FOR SYSTEMS WITH NUCLEAR HANKEL OPERATORS*

CHRIS GUIVER[†] AND MARK R. OPMEER[‡]

Abstract. We prove the H-infinity error bounds for Lyapunov balanced truncation and for optimal Hankel norm approximation under the assumption that the Hankel operator is nuclear. This is an improvement of the result from Glover, Curtain, and Partington [*SIAM J. Control Optim.*, 26 (1998), pp. 863–898], where additional assumptions were made. The proof is based on convergence of the Schmidt pairs of the Hankel operator in a Sobolev space. We also give an application of this convergence theory to a numerical algorithm for model reduction by balanced truncation.

Key words. infinite-dimensional system, model reduction, Hankel operator, realization, balanced realization, optimal Hankel norm approximation

AMS subject classifications. 93C25, 93B15, 41A65, 47B35, 93B28, 41A20, 47N70

DOI. 10.1137/110846981

1. Introduction. Approximation of a transfer function by truncation of a balanced state-space realization was first suggested for rational functions by Moore in [21]. A (Lyapunov) balanced realization is a realization where the controllability and observability Gramians are equal. The importance of balanced truncation relies on an explicit H^∞ bound on the difference of the transfer functions established independently by Enns [10] and Glover [11],

$$(1.1) \quad \|G - G_n\|_{H^\infty} \leq 2 \sum_{k=n+1}^N \sigma_k.$$

In the above inequality σ_k are the distinct singular values of the Hankel operator associated with G , of which there are N and n is the number kept in the reduced order system obtained by balanced truncation G_n . When all the singular values are simple, N and n are the dimensions of minimal state-space realizations of G and G_n , respectively.

There have been many extensions to the concepts introduced in [21] including those pertaining to bounded real systems [25], positive real systems [9], descriptor (or differential-algebraic or singular) systems [32], and behavioral systems [18].

In this article we are interested in the case where the transfer function is non-rational. In this case any realization must have an infinite-dimensional state space and the existence of balanced realizations is nontrivial. A special case was treated in Curtain and Glover [6] and the general discrete time case was proved by Young [41]. This was subsequently converted to general continuous-time systems by Ober and Montgomery-Smith [24]. In Glover, Curtain, and Partington [13] balanced truncations and the H^∞ error bound (with $N = \infty$ in (1.1)) were extended to a class

*Received by the editors September 6, 2011; accepted for publication (in revised form) February 25, 2014; published electronically April 29, 2014.

<http://www.siam.org/journals/sicon/52-2/84698.html>

[†]Environment & Sustainability Institute, College of Engineering, Mathematics and Physical Sciences, University of Exeter, Penryn Campus, TR10 9FE, UK (c.guiver@ex.ac.uk).

[‡]Department of Mathematical Sciences, University of Bath, Claverton Down, Bath, BA2 7AY, UK (m.opmeer@maths.bath.ac.uk).

of infinite-dimensional continuous-time systems. Balanced truncation for infinite-dimensional discrete time systems was later established by Bonnet [2, 3], where the corresponding H^∞ error bound was also proven.

The error bound (1.1) (with $N = \infty$) obviously holds when the right side is infinite, and so there is only something to prove for systems such that the series in (1.1) is finite (for some or equivalently for all n), which is certainly the case for systems that have a nuclear (or trace class) Hankel operator. In proving the error bound (1.1) (with $N = \infty$), the authors of [13] made further assumptions that are unnecessarily restrictive.

Therefore, in this paper we consider realization and approximation of systems with a nuclear Hankel operator and, contrary to [13], *we impose no further restrictions*. The question of which systems have nuclear Hankel operators has been addressed in, for example, Curtain and Sasane [7] and Opmeer [27]. As a byproduct, we also obtain a similar generalization of the H^∞ error bound for optimal Hankel norm approximations that was proven for the less general case in [13] and we obtain a convergence result that is of interest for numerical approximations of balanced truncations of irrational transfer functions.

The main technical novelties in our approach are that we consider realizations on L^1 rather than L^2 , as treated in [13], and that we also consider convergence of the Schmidt vectors of a Hankel operator in a Sobolev space. The article is organized as follows. Formal statements of the main results are given in section 2. The subsequent sections collect the material required for proving these results. Section 3 briefly describes Hankel operators on L^2 and gathers some known results in the notation used here. In section 4 we consider the convergence of the Schmidt pairs of such Hankel operators; this forms a crucial ingredient in the proofs of our main results. Section 5 focuses on the realization of G and the reduced order G_n . The proof of the error bound (1.1) (with $N = \infty$) is contained in section 6.

2. Summary of main results. Here we describe the main results of the paper. We do not provide all the details at this stage, but refer the reader to the relevant sections as appropriate. Our starting point is a nuclear (also known as trace class) Hankel operator

$$H : L^2(\mathbb{R}^+; \mathcal{U}) \rightarrow L^2(\mathbb{R}^+; \mathcal{Y}),$$

where $\mathbb{R}^+ = [0, \infty)$ and \mathcal{U} and \mathcal{Y} are the input and output spaces, respectively, which are always assumed finite dimensional. A key result from section 3 is a consequence of the nuclearity of H .

PROPOSITION 2.1. *If H is a nuclear Hankel operator then H is necessarily given by*

$$(2.1) \quad (Hf)(t) = \int_{\mathbb{R}^+} h(t+s)f(s) ds, \quad f \in L^2(\mathbb{R}^+; \mathcal{U}), \quad a.a. t \geq 0,$$

and the impulse response h is $L^1(\mathbb{R}^+; B(\mathcal{U}, \mathcal{Y}))$.

We direct the reader’s attention to section 4 where the definitions of Schmidt pairs and singular values of an operator are recapped. For now we recall that the (distinct) singular values $(\sigma_i)_{i \in \mathbb{N}}$ of a compact operator $T : L^2 \rightarrow L^2$ are the square roots of the countably many eigenvalues of T^*T , ordered such that

$$\sigma_1 > \sigma_2 > \dots \geq 0,$$

each with (geometric) multiplicity $p_i \in \mathbb{N}$. The Schmidt pairs $(v_{i,k}, w_{i,k})$ are the eigenvectors of T^*T and TT^* , respectively, corresponding to the eigenvalue σ_i^2 .

The transfer function G associated with a nuclear Hankel operator H is the Laplace transform of the impulse response h that appears in (2.1). Such a G always belongs to the Hardy space $H^\infty(\mathbb{C}_0^+; B(\mathcal{U}, \mathcal{Y}))$ of bounded, analytic $B(\mathcal{U}, \mathcal{Y})$ -valued functions on the open right-half complex plane \mathbb{C}_0^+ , and furthermore is regular with feedthrough zero in the terminology of Weiss [40].

A triple of operators (A, B, C) with suitable properties is called a *realization* of G if $G(s) = C(sI - A)^{-1}B$ on some right half-plane. Realizations are never unique. We choose the exactly observable shift realization (on L^1) of G , described in section 5, to define the balanced truncations. These truncated systems are defined using the Schmidt pairs of the Hankel operator. A key property is that the Schmidt vectors $w_{i,k}$ all belong to the Sobolev space $W^{1,1}(\mathbb{R}^+; \mathcal{Y})$, which is the domain of A for this particular realization.

DEFINITION 2.2. *Let (A, B, C) denote the generating operators of the exactly observable shift realization on $L^1(\mathbb{R}^+; \mathcal{Y})$ of a nuclear Hankel operator. For fixed $n \in \mathbb{N}$, we define a truncated state space \mathcal{X}_n as the linear span of Schmidt vectors of the Hankel operator by setting*

$$\mathcal{X}_n := \text{span} \{w_{i,k} \mid 1 \leq i \leq n, 1 \leq k \leq p_i\},$$

which is a closed subspace of L^1, L^2 , and $W^{1,1}(\mathbb{R}^+; \mathcal{Y})$. We define the truncated operators

$$(2.2) \quad A_n := \mathcal{P}_n A|_{\mathcal{X}_n}, \quad B_n := \mathcal{P}_n B, \quad C_n := C|_{\mathcal{X}_n},$$

where $\mathcal{P}_n : L^1(\mathbb{R}^+, \mathcal{Y}) \rightarrow \mathcal{X}_n$ is the projection onto \mathcal{X}_n defined by

$$\mathcal{P}_n x = \sum_{i=1}^n \sum_{k=1}^{p_i} \langle w_{i,k}, x \rangle_{L^2} w_{i,k} \quad \forall x \in L^1(\mathbb{R}^+; \mathcal{Y}).$$

Note that throughout this work we use the notation $\langle \cdot, \cdot \rangle_{L^2}$ to denote both the L^2 inner product and the sesquilinear form

$$(2.3) \quad \langle f, g \rangle_{L^2} = \int_{\mathbb{R}^+} \langle f(s), g(s) \rangle_{\mathcal{Z}} ds \quad \forall f \in L^\infty(\mathbb{R}^+; \mathcal{Z}), \forall g \in L^1(\mathbb{R}^+; \mathcal{Z}),$$

which is finite by the Hölder inequality. We will refer to (2.3) as the duality product between L^1 and L^∞ . Note that $W^{1,1}$ is continuously embedded in L^∞ .

The operators in (2.2) generate a finite-dimensional linear system on $(\mathcal{U}, \mathcal{X}_n, \mathcal{Y})$, which we sometimes denote by $\begin{bmatrix} A_n & B_n \\ C_n & 0 \end{bmatrix}$, called the balanced truncation. The function

$$s \mapsto G_n(s) := C_n(sI - A_n)^{-1}B_n,$$

defined on some right half-plane, is called the reduced order transfer function obtained by balanced truncation.

Definition 2.2 is consistent with earlier definitions of balanced truncation in the literature and in sections 4 and 5 we demonstrate that this definition is well-defined. We now state the main results.

THEOREM 2.3. *Let G denote a transfer function with nuclear Hankel operator. Then for any positive integer n*

$$(2.4) \quad \|G - G_n\|_{H^\infty} \leq 2 \sum_{k=n+1}^{\infty} \sigma_k,$$

where $(\sigma_k)_{k \in \mathbb{N}}$ is the sequence of distinct Hankel singular values and G_n is the reduced order transfer function obtained by balanced truncation.

Proof. Theorem 2.3 is proven in section 6. \square

As in [13], the proof of Theorem 2.3 utilizes an intermediate approximation. The specific intermediate approximation used is an abstract functional analytic construct, but the general result used, here formulated as Theorem 2.4, is also relevant for numerical approximations. In typical applications, h would be the impulse response of some controlled partial differential equation specified by (A, B, C) and h_m would be the impulse response of some semidiscretization (A_m, B_m, C_m) of (A, B, C) . The assumption that $h_m \rightarrow h$ in L^1 as $m \rightarrow \infty$ made in Theorem 2.4 is typical in such situations (see, for example, Morris [22]). The conclusion of Theorem 2.4 is that the object G_n^m , which can be computed explicitly, converges (along a subsequence) to G_n , the sought after object. See Singler [36] for further numerical results in this direction.

THEOREM 2.4. *Let G denote a transfer function whose Hankel operator is given by (2.1) with L^1 kernel h . For $n \in \mathbb{N}$, let G_n denote the reduced order transfer function obtained by balanced truncation, suppose that $(h_m)_{m \in \mathbb{N}}$ is such that*

$$h_m \xrightarrow{L^1} h, \quad \text{as } m \rightarrow \infty,$$

and let $(G^m)_{m \in \mathbb{N}}$ denote the corresponding sequence of transfer functions. If $(G_n^m)_{m \in \mathbb{N}}$ denotes the sequence of reduced order transfer functions obtained from G^m by balanced truncation, then there exists a subsequence $(\tau(m))_{m \in \mathbb{N}}$ such that

$$(2.5) \quad G_n^{\tau(m)} \xrightarrow{H^\infty} G_n, \quad \text{as } m \rightarrow \infty.$$

Proof. See Proposition 5.10 for the proof. \square

Remark 2.5. If all the Hankel singular values in Theorem 2.4 are simple then the convergence (2.5) does not require a subsequence.

2.1. The Glover, Curtain, and Partington [13] assumptions. Versions of Theorems 2.3 and 2.4 are proven in [13] under stronger assumptions on the original system. Specifically, their assumptions are

- (1) the Hankel operator $H : L^2(\mathbb{R}^+; \mathcal{U}) \rightarrow L^2(\mathbb{R}^+; \mathcal{Y})$ is of the form (2.1) with kernel $h \in L^1(\mathbb{R}^+; B(\mathcal{U}, \mathcal{Y}))$;
- (2) $h \in L^2(\mathbb{R}^+; B(\mathcal{U}, \mathcal{Y}))$;
- (3) the kernel h is real, or if h is complex then the derivative \dot{h} exists and \dot{h} is the kernel of a bounded Hankel operator;
- (4) the singular values of the Hankel operator are simple;
- (5) the Hankel operator H is nuclear.

We give an example of a physical system where assumption (2) fails and hence the results of [13] do not apply.

Example 2.6. Consider the following heat equation in one dimension on the unit interval

$$(2.6a) \quad w_t(t, x) = w_{xx}(t, x) \quad \forall t \geq 0, x \in [0, 1],$$

where subscripts denote partial derivatives with respect to that variable. We impose zero initial temperature profile, $w(0, x) = 0$, for all $x \in [0, 1]$, Neumann control, and Dirichlet observation at the left end

$$(2.6b) \quad \begin{aligned} u(t) &:= w_x(t, 0), \\ y(t) &:= w(t, 0) \end{aligned} \quad \forall t \geq 0,$$

so that the input and output spaces are one dimensional and there is a Dirichlet boundary condition at the right end

$$(2.6c) \quad w(t, 1) = 0 \quad \forall t \geq 0.$$

From [27, section 3, Theorem 3] it follows that the Hankel operator of (2.6) belongs to the Schatten class S_p for all $p > 0$, in particular, implying that the Hankel operator is nuclear. To find the transfer function G in this example we take the Laplace transform of (2.6) and solve the resulting ODE, which can be justified by Curtain and Zwart [8, Examples 4.3.11, 4.3.12]. Some elementary calculations give

$$(2.7) \quad G(s) = \frac{-1}{\sqrt{s}} \tanh(\sqrt{s}), \quad \text{Re } s > 0,$$

where we take the (unique) square root \sqrt{s} with argument in $(-\frac{\pi}{2}, \frac{\pi}{2})$, so that $\text{Re}\sqrt{s} > 0$.

We claim that $G \notin H^2(\mathbb{C}_0^+)$ which, as the Laplace transform is an isomorphism between $L^2(\mathbb{R}^+)$ and $H^2(\mathbb{C}_0^+)$, occurs if and only if $h \notin L^2(\mathbb{R}^+)$, where as usual h is the impulse response. To establish the claim note that along the line $\{\omega(1+i) : \omega \geq 0\}$ we have

$$\tanh(\omega(1+i)) = \frac{\tanh(\omega) + i \tan(\omega)}{1 + i \tanh(\omega) \tan(\omega)},$$

and so

$$(2.8) \quad |\tanh(\omega(1+i))|^2 = \frac{\tanh^2(\omega) + \tan^2(\omega)}{1 + \tanh^2(\omega) \tan^2(\omega)} \rightarrow 1 \quad \text{as } \omega \rightarrow \infty.$$

Therefore

$$(2.9) \quad \begin{aligned} \|G\|_{H^2}^2 &= \int_{\mathbb{R}} \left| \frac{-1}{\sqrt{\omega i}} \tanh(\sqrt{\omega i}) \right|^2 d\omega = \int_{\mathbb{R}} \frac{1}{|\omega|} \left| \tanh(\sqrt{\omega i}) \right|^2 d\omega, \\ &\geq \int_{\mathbb{R}^+} \frac{1}{\omega} \left| \tanh\left(\sqrt{\omega} \left(\frac{1}{\sqrt{2}} + \frac{i}{\sqrt{2}}\right)\right) \right|^2 d\omega \geq \int_C^\infty \frac{1}{2\omega} d\omega \end{aligned}$$

by (2.8), for some $C > 0$ sufficiently large. The second integral in (2.9) is infinite so that $G \notin H^2$ and hence $h \notin L^2$, that is, assumption (2) from [13] does not hold.

In this work we seek to carefully describe where the assumptions (1)–(5) are used in [13]. Proposition 2.1 above shows that nuclearity of H implies the integral operator form, that is, (5) implies (1). In sections 4.1 and 5.4 we explain how assumptions (2) and (3) can be avoided. It was already remarked in [13] that assumption (4) is unnecessary, but that its inclusion does make the notation in the proofs simpler.

2.2. Optimal Hankel norm approximations. We comment briefly on the optimal Hankel norm approximations of [13, section 6]. We refer the reader to that article and the references therein for more details. The following result is based on [13, Theorem 6.4] and the proof appeals to Theorem 2.3 instead of the corresponding result on balanced truncations from [13]. As such, the extra assumptions (1)–(4) from section 2.1 are unnecessary for optimal Hankel norm approximations as well. We comment that, unlike the H^∞ error bound (2.4), in what follows the multiplicities of the singular values *do* play a role. We recall precise definitions of these terms in Definition 4.2 and Remark 4.3.

THEOREM 2.7. *Let G denote a transfer function with nuclear Hankel operator and let $(\sigma_n)_{n \in \mathbb{N}}$ and $(p_n)_{n \in \mathbb{N}}$ denote the sequences of distinct Hankel singular values*

and their multiplicities, respectively. For each integer k there exists a transfer function $\hat{G}_k \in H^\infty$ of MacMillan degree $r_k := \sum_{i=1}^k p_i$ such that

$$\|G - \hat{G}_k\|_H = \sigma_{k+1},$$

where $\|F\|_H$ denotes the Hankel norm of the Hankel operator corresponding to the transfer function F . Thus \hat{G}_k is an optimal Hankel norm approximation for G . Moreover, there exists a constant $D_0 \in B(\mathcal{U}, \mathcal{Y})$ such that

$$\sigma_{k+1} \leq \|G - \hat{G}_k - D_0\|_{H^\infty} \leq \sigma_{k+1} + \sum_{j=k+2}^\infty p_j \sigma_j.$$

There exists $\tilde{G}_k \in H^\infty$ of MacMillan degree r_k such that

$$\sigma_{k+1} \leq \|G - \tilde{G}_k\|_{H^\infty} \leq \sum_{j=k+1}^\infty \sigma_j,$$

but \tilde{G}_k is not an optimal Hankel norm approximation of G in general.

Proof. The result mirrors [13, Theorem 6.4], where the authors (chiefly) extend the work of Glover [11, 12], but restrict to the case where all the singular values are simple. Crucially, [13, Theorem 5.1] is replaced by Theorem 2.3, where the stronger assumptions are not required. For nonsimple singular values, the treatment of optimal Hankel norm approximations for rational transfer functions by Zhou, Doyle, and Glover [42, section 8.3] can be extended by arguing in the same spirit as is found in [13, section 6]. \square

3. Hankel operators between $L^2(\mathbb{R}^+; \mathcal{Z})$ spaces. There is a large literature on Hankel operators; see, for example, [30], [23], [28], and [29], with unfortunately several different conventions used. The purpose of this section is to present a known result in the convention used here. Specifically, Proposition 3.4, which states that a nuclear Hankel operator $L^2(\mathbb{R}^+; \mathcal{U}) \rightarrow L^2(\mathbb{R}^+; \mathcal{Y})$ is necessarily an integral operator of the form (2.1) with L^1 kernel.

We start by recalling the definition of a Hankel operator in an abstract Hilbert space setting based on shift operators. The following definition is taken from Rosenblum and Rovnyak [33, p. 1].

DEFINITION 3.1. *Let \mathcal{H} denote a Hilbert space. An operator $S \in B(\mathcal{H})$ is a shift on \mathcal{H} if S is an isometry and $(S^*)^n$ converges strongly to zero as n tends to infinity. We define a (bounded) S -Hankel operator H on \mathcal{H} as a bounded operator which satisfies*

$$S^*H = HS.$$

It is possible to define Hankel operators between two different spaces. If \mathcal{L} is another Hilbert space, the operator $H \in B(\mathcal{H}, \mathcal{L})$ is (S_1, S_2) -Hankel if there exist shift operators $S_1 \in B(\mathcal{H})$, $S_2 \in B(\mathcal{L})$ such that

$$S_2^*H = HS_1.$$

We collect results for Hankel operators from $L^2(\mathbb{R}^+, \mathcal{Z}_1)$ to $L^2(\mathbb{R}^+, \mathcal{Z}_2)$ corresponding to the right-shift semigroup described below, where $\mathcal{Z}_1, \mathcal{Z}_2$ are arbitrary Hilbert spaces.

DEFINITION 3.2. For a Hilbert space \mathcal{X} , define the family of right-shift operators $(\mathbb{S}_{\mathcal{X}}(\tau))_{\tau \geq 0}$ on $L^2(\mathbb{R}^+; \mathcal{X})$ by

$$(\mathbb{S}_{\mathcal{X}}(\tau)f)(t) = \begin{cases} f(t - \tau) & t \geq \tau, \\ 0 & t < \tau. \end{cases}$$

The adjoint operators are the left-shift operators given by

$$(\mathbb{S}_{\mathcal{X}}^*(\tau)f)(t) = f(t + \tau) \quad \forall t, \tau \geq 0.$$

DEFINITION 3.3. The operator H is $(\mathbb{S}_{\mathcal{X}_1}, \mathbb{S}_{\mathcal{X}_2})$ -Hankel if it is $(\mathbb{S}_{\mathcal{X}_1}(\tau), \mathbb{S}_{\mathcal{X}_2}(\tau))$ -Hankel for every $\tau \geq 0$, so that

$$\mathbb{S}_{\mathcal{X}_2}^*(\tau)H = H\mathbb{S}_{\mathcal{X}_1}(\tau) \quad \forall \tau \geq 0.$$

From now on all Hankel operators considered are $(\mathbb{S}_{\mathcal{X}_1}, \mathbb{S}_{\mathcal{X}_2})$ -Hankel (and bounded), and so we shall omit the prefix $(\mathbb{S}_{\mathcal{X}_1}, \mathbb{S}_{\mathcal{X}_2})$. Transfer functions of a nuclear Hankel operator have atomic decompositions given by Coifman and Rochberg [5] and the following result which uses these decompositions is based on Partington [29, Corollary 7.9].

PROPOSITION 3.4. Let \mathcal{X}_1 and \mathcal{X}_2 denote finite-dimensional Hilbert spaces. A Hankel operator $H : L^2(\mathbb{R}^+; \mathcal{X}_1) \rightarrow L^2(\mathbb{R}^+; \mathcal{X}_2)$ is nuclear if and only if H is an integral operator given by

$$(3.1) \quad (Hf)(t) = \int_0^\infty h(t+s)f(s) ds \quad \forall f \in L^2(\mathbb{R}^+; \mathcal{X}_1), \text{ a.a } t \geq 0,$$

with $h \in L^1(\mathbb{R}^+; B(\mathcal{X}_1, \mathcal{X}_2))$ satisfying

$$(3.2) \quad h(t) := \sum_{n \in \mathbb{N}} \lambda_n (\operatorname{Re} a_n) e^{a_n t}, \quad t > 0,$$

for sequences $(\lambda_n)_{n \in \mathbb{N}} \in \ell^1(B(\mathcal{X}_1, \mathcal{X}_2))$ and $(a_n)_{n \in \mathbb{N}} \subseteq \mathbb{C}_0^-$. The function

$$(3.3) \quad G(s) = \sum_{n \in \mathbb{N}} \lambda_n \frac{\operatorname{Re} a_n}{s - a_n}, \quad \operatorname{Re} s > 0,$$

is the Laplace transform of h and $G \in H^\infty(\mathbb{C}_0^+, B(\mathcal{X}_1, \mathcal{X}_2))$ and is the transfer function of the system with Hankel operator H . Furthermore, if h_p denotes the p th partial sum of (3.2) with corresponding Hankel operator H_p then H_p converges to H in nuclear norm as $p \rightarrow \infty$.

Remark 3.5. The transfer function G of a Hankel operator is only determined by the Hankel operator up to the addition of a constant operator. For the transfer function in (3.3) that constant is fixed by the condition

$$\lim_{\substack{s \in \mathbb{R} \\ s \rightarrow \infty}} G(s) = 0.$$

We see that the transfer function is regular with zero feedthrough [40].

Remark 3.6. The Coifman and Rochberg atomic decomposition (3.3) can equivalently be expressed as any nuclear Hankel operator is the (possibly infinite, but absolutely convergent) sum of rank one Hankel operators. A discrete time version

of this equivalence is given in Peller [30, p. 237]. We note that the transfer function of a Hankel operator is rational if and only if the corresponding Hankel operator is finite rank; this is Kronecker’s theorem; see, for example, [29, Corollary 4.9]. Every finite-rank Hankel operator is nuclear, and thus every rational transfer function admits a decomposition of the form (3.3), although in general this sum is infinite. The latter case occurs when the rational transfer function has a repeated pole, such as $s \mapsto \frac{1}{(s+1)^2}$. Computing a Coifman and Rochberg atomic decomposition exactly in general is, as known to the authors, intractable.

Remark 3.7.

(i) In [13, Theorem 2.1] the following bound is proven:

$$(3.4) \quad \|h\|_{L^1} \leq 2\|H\|_N$$

for a nuclear Hankel operator H given by (2.1). However, the proof of (3.4) in [13] uses the fact that $h \in L^1$. It seems to have been missed that the atomic decomposition (3.2) implies that $h \in L^1$. That $h \in L^1$ can be seen from the monotone convergence theorem applied to the partial sums

$$(0, \infty) \ni t \mapsto \sum_{n=1}^M \|\lambda_n(\operatorname{Re} a_n)e^{a_n t}\| \in L^1.$$

(ii) The inequality (3.4) demonstrates that $h \in L^1$ is necessary for the Hankel operator H to be nuclear. However, note further that $h \in L^1$ is *not* sufficient for nuclearity of the Hankel operator given by (3.1). A counterexample is contained in Glover, Lam, and Partington [14, Example 2.3]. There they show that although the kernel h given by

$$[0, \infty) \ni t \mapsto h(t) := e^{-t}\chi_{[1, \infty)}(t),$$

where χ_J is the indicator (also called the characteristic) function on $J \subseteq \mathbb{R}^+$, clearly belongs to L^1 , the Hankel operator (3.1) with kernel h is not nuclear.

4. Convergence of Schmidt pairs of integral Hankel operators. In this section we describe properties of the Schmidt pairs of a Hankel operator given by (2.1) with L^1 kernel, which by Proposition 3.4 includes nuclear Hankel operators. We derive a convergence result, Theorem 4.4, for the Schmidt pairs when the kernel is approximated in L^1 that is crucial in proving our main results. We first collect some elementary facts regarding Sobolev spaces.

Remark 4.1. For \mathcal{H} a Hilbert space, $m \in \mathbb{N}$, and $1 \leq p < \infty$, the Sobolev space $W^{m,p} = W^{m,p}(\mathbb{R}^+; \mathcal{H})$ is defined as $f \in L^p(\mathbb{R}^+; \mathcal{H})$ such that all m (distributional) derivatives $f^{(1)}, \dots, f^{(m)}$ belong to L^p as well. The norm on $W^{m,p}$ is given by

$$\|f\|_{m,p} = \left(\|f\|_p + \sum_{j=1}^m \|f^{(j)}\|_p \right)^{\frac{1}{p}}.$$

We are mostly interested in $W^{1,1}$ and $W^{1,2}$, and note the following facts:

(i) $W^{1,1}$ and $W^{1,2}$ are both continuously embedded into the space of continuous functions on \mathbb{R}^+ with the supremum norm,

$$(4.1) \quad v \in W^{1,1} \Rightarrow \|v\|_\infty \leq \|v\|_{1,1} \quad \text{and} \quad w \in W^{1,2} \Rightarrow \|w\|_\infty \leq \sqrt{2}\|w\|_{1,2}.$$

Both of the estimates in (4.1) are readily established by writing for $x \in \mathbb{R}^+$,

$$v(x) = - \int_x^\infty v'(t) dt \quad \text{and} \quad \|w(x)\|^2 = - \int_x^\infty \frac{d}{dt} \langle w(t), w(t) \rangle dt,$$

respectively. Consequently, when arguing with $W^{1,1}$ and $W^{1,2}$ functions we shall always choose a continuous representative (so that point evaluations are well-defined).

- (ii) $W^{1,1}$ is continuously embedded in L^2 , which follows from the Hölder inequality and (4.1), namely,

$$v \in W^{1,1} \quad \Rightarrow \quad \|v\|_2^2 \leq \|v\|_1 \cdot \|v\|_\infty \leq \|v\|_{1,1}^2.$$

We recall the definition of singular values.

DEFINITION 4.2. *Let $\mathcal{B}_1, \mathcal{B}_2$ denote Banach spaces and $T \in B(\mathcal{B}_1, \mathcal{B}_2)$. For $k \in \mathbb{N}$, the k th singular value (also called s -value or approximation number) of T , denoted s_k , is defined as*

$$s_k := \inf \{ \|T - T_k\| : \text{rank } T_k \leq k - 1 \}.$$

We define the k th distinct singular value σ_k of T as follows. If $s_1 = s_2 = \dots = s_{p_1}$ and $s_{p_1} > s_{p_1+1}$, for some $p_1 \in \mathbb{N}$, then we set

$$\sigma_1 = s_1 = s_2 = \dots = s_{p_1}, \quad \sigma_2 = s_{p_1+1} = \dots,$$

and so on. As such, the k th distinct singular value σ_k has multiplicity $p_k \in \mathbb{N}$ and satisfies $\sigma_k > \sigma_{k+1}$, although σ_k now need not necessarily be the distance of T to rank $k - 1$ operators. Using this convention the operator T is nuclear if

$$\sum_{k \in \mathbb{N}} s_k = \sum_{k \in \mathbb{N}} p_k \sigma_k < \infty.$$

We comment that if all the singular values are simple, then the sequences of singular values and distinct singular values coincide.

If \mathcal{H}_1 and \mathcal{H}_2 are Hilbert spaces and $T \in B(\mathcal{H}_1, \mathcal{H}_2)$ is compact, then the distinct singular values are precisely the square roots of the countably many eigenvalues of T^*T , indexed in decreasing order. In this instance we denote by p_i the (finite) (geometric) multiplicity of the eigenvalue σ_i^2 . The Schmidt pairs $(v_{i,k}, w_{i,k})$ of T are eigenvectors of T^*T and TT^* , respectively, corresponding to a singular value σ_i . The Schmidt pairs can be chosen to satisfy

$$\left. \begin{aligned} w_{i,k} \in \mathcal{H}_2, \quad T v_{i,k} &= \sigma_i w_{i,k}, \\ v_{i,k} \in \mathcal{H}_1, \quad T^* w_{i,k} &= \sigma_i v_{i,k}, \end{aligned} \right\} \quad \forall i \in \mathbb{N}, \quad 1 \leq k \leq p_i,$$

and are always chosen orthonormal.

Remark 4.3. Singular values of compact operators between Hilbert spaces are often *defined* as the square roots of the eigenvalues, indexed in decreasing order, as in Lax [20, p. 330], for example. By Gohberg, Goldberg, and Kaashoek [15, Theorem VI. 1.5] the definition of singular value found in [20] is equivalent to that in Definition 4.2.

The key assumption of this section is the following.

- (A) The operator $H : L^2(\mathbb{R}^+; \mathcal{U}) \rightarrow L^2(\mathbb{R}^+; \mathcal{Y})$ is a Hankel¹ operator given by (2.1) with kernel $h \in L^1(\mathbb{R}^+; B(\mathcal{U}, \mathcal{Y}))$. The input space \mathcal{U} and output space \mathcal{Y} are finite-dimensional Hilbert spaces.

¹More precisely $(\mathbb{S}_{\mathcal{U}}, \mathbb{S}_{\mathcal{Y}})$ -Hankel in the terminology of section 3.

Note that under assumption (A) it follows from [13, Appendix 1] that H is a compact operator $L^2(\mathbb{R}^+; \mathcal{U}) \rightarrow L^2(\mathbb{R}^+; \mathcal{Y})$, and so the singular values of H are precisely the square roots of the countably many eigenvalues of H^*H , ordered in decreasing magnitude. The following theorem shows that for a sequence of integral operators of the form (2.1), convergence of the impulse response implies convergence of the Schmidt vectors of the operators in L^2 and also in the Sobolev space $W^{1,1}$.

THEOREM 4.4. *Assume that H satisfies assumption (A) and let $(\sigma_k)_{k \in \mathbb{N}}$ denote the distinct singular values of H , each with multiplicity $p_k \in \mathbb{N}$. Let $(h_m)_{m \in \mathbb{N}}$ denote any sequence of kernels approximating h in the sense that*

$$h_m \xrightarrow{L^1} h, \quad \text{as } m \rightarrow \infty.$$

Define $(H_m)_{m \in \mathbb{N}}$ as the sequence of Hankel operators $L^2(\mathbb{R}^+; \mathcal{U}) \rightarrow L^2(\mathbb{R}^+; \mathcal{Y})$ given by the integral operators

$$(4.2) \quad (H_m f)(t) := \int_0^\infty h_m(t+s)f(s) \, ds \quad \forall f \in L^2(\mathbb{R}^+; \mathcal{U}), \quad \text{a.a. } t \geq 0.$$

Let $(\sigma_i^{(m)})_{i \in \mathbb{N}}$ denote the distinct singular values of H_m , also ordered in decreasing magnitude, each with multiplicity $p_i^{(m)} \in \mathbb{N}$. Then for all $k \in \mathbb{N}$ there exists $l_k \in \mathbb{N}$ such that with $l_0 := 0$,

$$(4.3) \quad \text{and } \left. \begin{aligned} &\sigma_i^{(m)} \rightarrow \sigma_k \quad \text{for } i \in \{l_{k-1} + 1, \dots, l_k\} \\ &\sum_{i=l_{k-1}+1}^{l_k} p_i^{(m)} \rightarrow p_k, \end{aligned} \right\} \text{ as } m \rightarrow \infty.$$

Choose orthonormal Schmidt pairs of H_m denoted by $(v_{i,r}^{(m)}, w_{i,r}^{(m)})$, where

$$r \in \{1, 2, \dots, p_i^{(m)}\}.$$

Then

- (i) the Schmidt pairs satisfy $v_{i,r}^{(m)} \in W^{1,1}(\mathbb{R}^+; \mathcal{U}) \subseteq L^2$, $w_{i,r}^{(m)} \in W^{1,1}(\mathbb{R}^+; \mathcal{Y}) \subseteq L^2$, for all $i, m \in \mathbb{N}$ and all $r \in \{1, 2, \dots, p_i^{(m)}\}$;
- (ii) there exists a subsequence $(m_j)_{j \in \mathbb{N}}$ along which for each $k \in \mathbb{N}$ there exists $i \in \mathbb{N}$, $q \in \{1, 2, \dots, p_k\}$, and r such that

$$(4.4) \quad \left. \begin{aligned} &v_{i,r}^{(m_j)} \xrightarrow{W^{1,1}} v_{k,q} \\ &w_{i,r}^{(m_j)} \xrightarrow{W^{1,1}} w_{k,q} \end{aligned} \right\} \text{ as } j \rightarrow \infty;$$

- (iii) the $(v_{k,q}, w_{k,q})$ are Schmidt pairs for H corresponding to σ_k and $(w_{k,q})_{\substack{1 \leq q \leq p_k \\ k \in \mathbb{N}}}$ form an orthonormal basis of eigenvectors in L^2 for HH^* and $(v_{k,q})_{\substack{1 \leq q \leq p_k \\ k \in \mathbb{N}}}$ for H^*H .

Remark 4.5.

- (i) The statement of Theorem 4.4 is notation heavy in order to account for the multiplicities of both the distinct singular values $(\sigma_k)_{k \in \mathbb{N}}$ of H and $(\sigma_k^{(m)})_{k \in \mathbb{N}}$ of H_m . The easiest case to understand is when H and H_m have simple singular values for every $m \in \mathbb{N}$, which is the case considered in [13]. The nonsimple case is conceptually similar, and we treat it for full generality, although the proofs become more complicated. Intuitively, what is important is that there is a subsequence $(m_j)_{j \in \mathbb{N}}$ along which every sequence $(v_{k,r}^{(m_j)})_{j \in \mathbb{N}}$ and $(w_{k,r}^{(m_j)})_{j \in \mathbb{N}}$ has a $W^{1,1}$ limit (which is also a limit in L^2), and we get

“enough” limits, in the sense that the limits of $(v_{k,r}^{(m_j)})_{j \in \mathbb{N}}$ and $(w_{k,r}^{(m_j)})_{j \in \mathbb{N}}$ form an orthonormal basis of eigenvectors for H^*H and HH^* , respectively.

- (ii) If in addition to the assumptions of Theorem 4.4, all the singular values of H are simple, then the convergence of Schmidt pairs in Theorem 4.4 does not require a subsequence. We do not give the details, but this follows from [13, Appendix 2, p. 896] and Lemma 4.7 below combined with [4, Exercise 5.5].
- (iii) Note that for this section we do not need to assume that our original Hankel operator H is nuclear, instead only that it satisfies assumption (A). Certainly, by Proposition 3.4, nuclearity of H is sufficient for (A) to hold.

The remainder of this section is dedicated to proving Theorem 4.4. We collect two required technical results, the first from Lax [19] and the second based on Chatelin [4, Theorem 5.10] and [4, Table 5.1].

LEMMA 4.6. *Let \mathcal{B} denote a Banach space on which is defined a continuous sesquilinear form $\langle \cdot, \cdot \rangle$ which induces a new norm under which the completion of \mathcal{B} is a Hilbert space \mathcal{H} . Suppose that T is a bounded operator on \mathcal{B} such that*

$$\langle Tx, y \rangle = \langle x, Ty \rangle \quad \forall x, y \in \mathcal{B}.$$

Then

- (1) T extends by continuity to an operator in $B(\mathcal{H})$;
- (2) the spectrum of T over \mathcal{H} is a subset of the spectrum of T over \mathcal{B} ;
- (3) the point spectrum of T over \mathcal{B} is contained in the point spectrum of T over \mathcal{H} and the eigenspace of T over \mathcal{B} with respect to an eigenvalue λ is the same as the eigenspace of T over \mathcal{H} with respect to the same eigenvalue.

LEMMA 4.7. *Suppose that $T, T_m : \mathcal{B} \rightarrow \mathcal{B}$ are compact operators on a Banach space \mathcal{B} such that $T_m \rightarrow T$ uniformly as $m \rightarrow \infty$. Let λ denote an eigenvalue of T and let $(\lambda_m)_{m \in \mathbb{N}}$ denote a sequence of eigenvalues of $(T_m)_{m \in \mathbb{N}}$. If $\lambda_m \rightarrow \lambda$ as $m \rightarrow \infty$, and $(v_m)_{m \in \mathbb{N}}$ is a uniformly bounded sequence of eigenvectors corresponding to λ_m , then there exists a subsequence $(v_{m_j})_{j \in \mathbb{N}}$ that converges to an eigenvector of T corresponding to λ .*

Note that Lemma 4.7 states that the limit of a uniformly bounded sequence of eigenvectors of $(T_m)_{m \in \mathbb{N}}$ is an eigenvector of T , but not that every eigenvector of T is obtained in this way. In the general case, this latter statement is not true (see, for example, [4, Example 5.5]). However, for self-adjoint operators on a Hilbert space it is true (a proof can be found, for instance, in [26, Lemma 10.19]). To establish Theorem 4.4 we show that, loosely speaking, the convergence in the weaker L^2 Hilbert space norm combined with continuous embedding of $W^{1,1}$ in L^2 gives convergence of the Schmidt pairs in the stronger $W^{1,1}$ Banach space norm. Lemmas 4.8, 4.9, and 4.10 contain the details.

LEMMA 4.8. *Let \mathcal{B} denote a Banach space continuously embedded into a Hilbert space \mathcal{H} , so that there exists a constant $C > 0$ such that*

$$(4.5) \quad \|v\|_{\mathcal{H}} \leq C \|v\|_{\mathcal{B}} \quad \forall v \in \mathcal{B}.$$

Let T, T_m denote compact operators on \mathcal{B} such that

$$(4.6) \quad T_m \rightarrow T \quad \text{uniformly on } \mathcal{B} \text{ as } m \rightarrow \infty.$$

Fix an eigenvalue λ of T with corresponding eigenvector v . Suppose that there exists a sequence $(\lambda_m)_{m \in \mathbb{N}}$, where λ_m is an eigenvalue of T_m , such that

$$(4.7) \quad \lambda_m \rightarrow \lambda, \quad \text{as } m \rightarrow \infty,$$

and also there exist eigenvectors $v^{(m)}$ of T_m corresponding to λ_m , which are orthonormal in \mathcal{H} , and such that

$$(4.8) \quad v^{(m)} \xrightarrow{\mathcal{H}} v, \quad \text{as } m \rightarrow \infty.$$

Then it follows that along a subsequence

$$(4.9) \quad v^{(m_j)} \xrightarrow{\mathcal{B}} v, \quad \text{as } j \rightarrow \infty.$$

Proof. Without loss of generality we may assume that the constant C in (4.5) is equal to one (else define an equivalent norm $C\|\cdot\|_{\mathcal{B}}$ on \mathcal{B} which induces the same topology on \mathcal{B}). Fix an eigenvalue λ and corresponding eigenvector $0 \neq v \in \mathcal{B}$ of T . By assumption there exists a sequence $(v_m)_{m \in \mathbb{N}}$ of eigenvectors of T_m that are orthonormal in \mathcal{H} and such that (4.8) holds. We seek to prove that the convergence in (4.9) holds as well. To that end for $m \in \mathbb{N}$ define

$$(4.10) \quad z^{(m)} := \frac{\|v\|_{\mathcal{B}}}{\|v^{(m)}\|_{\mathcal{B}}} v^{(m)} \quad \text{for which } \|z^{(m)}\|_{\mathcal{B}} = \|v\|_{\mathcal{B}} < \infty.$$

The sequence $(z^{(m)})_{m \in \mathbb{N}}$ satisfies the hypotheses of Lemma 4.7 and so there exists a subsequence (not relabeled) along which

$$(4.11) \quad z^{(m)} \xrightarrow{\mathcal{B}} \psi, \quad \text{as } m \rightarrow \infty,$$

with ψ an eigenvector of T . It remains to show that $\psi = v$. From (4.10) and (4.11) we see that

$$(4.12) \quad \|\psi\|_{\mathcal{B}} = \lim_{m \rightarrow \infty} \|z^{(m)}\|_{\mathcal{B}} = \|v\|_{\mathcal{B}} > 0,$$

and from the continuous embedding (4.5), convergence in (4.11) gives

$$(4.13) \quad z^{(m)} \xrightarrow{\mathcal{H}} \psi, \quad \text{as } m \rightarrow \infty.$$

We want to compare the convergence in (4.8) and (4.13), using the definition of $z^{(m)}$ in (4.10). If $\|v^{(m)}\|_{\mathcal{B}}$ is unbounded, then by (4.8) and (4.10) we see that

$$(4.14) \quad z^{(m)} \xrightarrow{\mathcal{H}} 0_{\mathcal{H}}, \quad \text{as } m \rightarrow \infty,$$

and so $\psi = 0_{\mathcal{H}}$ by uniqueness of limits. However, as $\mathcal{B} \subseteq \mathcal{H}$ are vector spaces they share the same zero element and thus we obtain the contradiction $0_{\mathcal{B}} \neq \psi = 0_{\mathcal{H}} = 0_{\mathcal{B}}$. Consequently, $\|v^{(m)}\|_{\mathcal{B}}$ is bounded and so has a convergent subsequence (not relabeled) with limit $B \geq 0$. From the continuous embedding (4.5)

$$1 = \|v^{(m)}\|_{\mathcal{H}} \leq \|v^{(m)}\|_{\mathcal{B}} \quad \forall m \in \mathbb{N},$$

it follows that $B \geq 1$. Combining now the definition (4.10) of $z^{(m)}$, the algebra of limits, and the convergence (4.8) we obtain

$$z^{(m)} \xrightarrow{\mathcal{H}} \frac{\|v\|_{\mathcal{B}} v}{B}, \quad \text{as } m \rightarrow \infty,$$

which when compared with (4.13) yields

$$(4.15) \quad \psi = \frac{\|v\|_{\mathcal{B}} v}{B} \quad \text{in } \mathcal{H}.$$

As both sides of (4.15) belong to \mathcal{B} , equality (4.15) and the injectivity of the inclusion $\mathcal{B} \hookrightarrow \mathcal{H}$ imply that

$$(4.16) \quad \psi = \frac{\|v\|_{\mathcal{B}}v}{B} \quad \text{in } \mathcal{B}.$$

Taking \mathcal{B} norms in (4.16) and using (4.12) gives $B = \|v\|_{\mathcal{B}}$ which when substituted back into (4.16) yields that $\psi = v$, as required. \square

We now turn attention to Hankel operators satisfying (A) and suitable approximations with a view to applying to Lemmas 4.6–4.8 and thus establishing the desired convergence of the Schmidt pairs.

LEMMA 4.9. *Let H denote a Hankel operator satisfying (A). Then H is a compact operator $L^2(\mathbb{R}^+; \mathcal{U}) \rightarrow L^2(\mathbb{R}^+; \mathcal{Y})$, $W^{1,1}(\mathbb{R}^+; \mathcal{U}) \rightarrow W^{1,1}(\mathbb{R}^+; \mathcal{Y})$, and every Schmidt pair (v, w) of H satisfies*

$$v \in W^{1,1}(\mathbb{R}^+; \mathcal{U}), \quad w \in W^{1,1}(\mathbb{R}^+; \mathcal{Y}).$$

The Hilbert space adjoint operator H^* satisfies (A) with $h(t)$ replaced by $h^*(t) = (h(t))^*$ and \mathcal{U} and \mathcal{Y} interchanged. Thus H^* is a compact operator $L^2(\mathbb{R}^+; \mathcal{Y}) \rightarrow L^2(\mathbb{R}^+; \mathcal{U})$, $W^{1,1}(\mathbb{R}^+; \mathcal{Y}) \rightarrow W^{1,1}(\mathbb{R}^+; \mathcal{U})$, and in both cases is defined by

$$(4.17) \quad (H^*f)(t) = \int_{\mathbb{R}^+} h^*(t+s)f(s) ds \quad \forall f \in L^2(\mathbb{R}^+; \mathcal{Y}).$$

Proof. It is proven in [13, Appendix 1, p. 895] that H satisfying (A) is a compact operator $L^1(\mathbb{R}^+; \mathcal{U}) \rightarrow L^1(\mathbb{R}^+; \mathcal{Y})$, $L^2(\mathbb{R}^+; \mathcal{U}) \rightarrow L^2(\mathbb{R}^+; \mathcal{Y})$. We abuse notation and use the symbol H to represent any of these maps. Let $\|H\|_1$ ($\|H\|_2$) denote the Hankel (operator) norm of a Hankel operator H satisfying (A) considered as an operator on L^1 (L^2). The key estimates

$$(4.18) \quad \|H\|_1, \|H\|_2 \leq \|h\|_1,$$

which we shall make frequent use of, are proven in [13, Appendix 1, p. 895].

We first prove that H is a bounded operator $W^{1,1}(\mathbb{R}^+; \mathcal{U}) \rightarrow W^{1,1}(\mathbb{R}^+; \mathcal{Y})$ and for this we need the following formula for the derivative of the Hankel operator given by (2.1):

$$(4.19) \quad (Hf)' = -hf(0) - H\dot{f} \quad \forall f \in W^{1,1}(\mathbb{R}^+; \mathcal{U}),$$

noting that we use both notations \dot{g} and g' to denote the first derivative of g . We recall Remark 4.1 in that we always choose continuous representatives of $W^{1,1}$ functions to understand point evaluations such as those in (4.19). The derivation of (4.19) is given in [13, Appendix 1]. For $f \in W^{1,1}$ we consider

$$(4.20) \quad \|Hf\|_{1,1} = \|Hf\|_1 + \|(Hf)'\|_1.$$

The first term on the right-hand side of (4.20) is clearly bounded by

$$(4.21) \quad \|H\|_1 \cdot \|f\|_1 \leq \|h\|_1 \cdot \|f\|_{1,1},$$

where we have used the bound (4.18). We now estimate the second term on the right-hand side of (4.20). From the formula for the derivative (4.19) we see that

$$(4.22) \quad \begin{aligned} \|(Hf)'\|_1 &\leq \|h(\cdot)f(0)\|_1 + \|H\dot{f}\|_1 \leq \|h\|_1 \cdot \|f(0)\|_{\mathcal{U}} + \|H\|_1 \cdot \|\dot{f}\|_1 \\ &\leq \|h\|_1 \cdot \|f\|_{\infty} + \|H\|_1 \cdot \|\dot{f}\|_1 \leq 2\|h\|_1 \cdot \|f\|_{1,1}, \end{aligned}$$

where we have used the estimates (4.18) and (4.1). Inserting (4.21) and (4.22) into (4.20) we obtain

$$(4.23) \quad \|Hf\|_{1,1} \leq 3\|h\|_1 \cdot \|f\|_{1,1},$$

and so $H : W^{1,1} \rightarrow W^{1,1}$ is well-defined and bounded. To prove compactness let $(f_n)_{n \in \mathbb{N}} \subseteq W^{1,1}$ denote a bounded sequence in $W^{1,1}$, which is therefore also a bounded sequence in L^1 . Since $H : L^1 \rightarrow L^1$ is compact, there exists a convergent and so Cauchy (in L^1) subsequence $(Hf_{\tau_1(n)})_{n \in \mathbb{N}} \subseteq L^1$. The sequence $(\dot{f}_{\tau_1(n)})_{n \in \mathbb{N}}$ is bounded in L^1 and so by the same argument, $(H\dot{f}_{\tau_1(n)})_{n \in \mathbb{N}}$ has a convergent and hence Cauchy subsequence denoted $(H\dot{f}_{\tau_2(n)})_{n \in \mathbb{N}}$. Observe that the trace map

$$T : W^{1,1}(\mathbb{R}^+; \mathcal{U}) \rightarrow \mathcal{U}, \quad Tu = u(0),$$

is bounded and finite rank and so is compact. Therefore there exists a subsequence of $(Tf_{\tau_2(n)} = f_{\tau_2(n)}(0))_{n \in \mathbb{N}}$, denoted by $(Tf_{\tau_3(n)})_{n \in \mathbb{N}}$, that is convergent in \mathcal{U} and so Cauchy in \mathcal{U} . We now compute for $m, n \in \mathbb{N}$,

$$(4.24) \quad \begin{aligned} \|Hf_{\tau_3(n)} - Hf_{\tau_3(m)}\|_{1,1} &= \|Hf_{\tau_3(n)} - Hf_{\tau_3(m)}\|_1 + \|(Hf_{\tau_3(n)} - Hf_{\tau_3(m)})'\|_1 \\ &\leq \|Hf_{\tau_3(n)} - Hf_{\tau_3(m)}\|_1 + \|H\dot{f}_{\tau_3(n)} - H\dot{f}_{\tau_3(m)}\|_1 \\ &\quad + \|h\|_1 \cdot \|f_{\tau_3(n)}(0) - f_{\tau_3(m)}(0)\|_{\mathcal{U}}, \end{aligned}$$

where we have used the formula (4.19) for the derivative of Hf_{τ_3} . By construction the right-hand side of (4.24) converges to zero. Thus, the sequence $(Hf_{\tau_3(n)})_{n \in \mathbb{N}}$ is Cauchy in $W^{1,1}$ and so convergent, completing the proof that $H : W^{1,1} \rightarrow W^{1,1}$ is compact.

We now focus our attention on the Hilbert space adjoint map H^* . First note that $h^* \in L^1(\mathbb{R}^+; B(\mathcal{Y}, \mathcal{U}))$ as

$$\|h^*\|_1 = \int_{\mathbb{R}^+} \|h^*(t)\|_{B(\mathcal{Y}, \mathcal{U})} dt = \int_{\mathbb{R}^+} \|h(t)\|_{B(\mathcal{U}, \mathcal{Y})} dt = \|h\|_1,$$

where $h^*(t) = (h(t))^*$. A short calculation shows that $H^* : L^2(\mathbb{R}^+; \mathcal{Y}) \rightarrow L^2(\mathbb{R}^+; \mathcal{U})$ is indeed given by (4.17) and so H^* certainly satisfies (A) with \mathcal{U} and \mathcal{Y} interchanged and h replaced by h^* . The operator H^* is compact $L^2(\mathbb{R}^+; \mathcal{Y}) \rightarrow L^2(\mathbb{R}^+; \mathcal{U})$ and $W^{1,1}(\mathbb{R}^+; \mathcal{Y}) \rightarrow W^{1,1}(\mathbb{R}^+; \mathcal{U})$ for the same reasons that H is. The claims for the Schmidt pairs (v, w) of H now follow from Lemma 4.6, with

$$(4.25) \quad \begin{aligned} \mathcal{B} = \mathcal{B}_{\mathcal{U}} = W^{1,1}(\mathbb{R}^+; \mathcal{U}), \quad \mathcal{H} = \mathcal{H}_{\mathcal{U}} = L^2(\mathbb{R}^+; \mathcal{U}), \quad T = H^*H, \\ \text{and } \mathcal{B} = \mathcal{B}_{\mathcal{Y}} = W^{1,1}(\mathbb{R}^+; \mathcal{Y}), \quad \mathcal{H} = \mathcal{H}_{\mathcal{Y}} = L^2(\mathbb{R}^+; \mathcal{Y}), \quad T = HH^* \end{aligned}$$

for v and w , respectively. In both cases we use that T is symmetric with respect to $\langle \cdot, \cdot \rangle_{L^2}$, that the closure of $W^{1,1}$ in L^2 is L^2 , and that as T is compact (in both \mathcal{B} and \mathcal{H}), $\sigma(T) = \sigma_p(T) \cup \{0\}$, where σ_p denotes the point spectrum. \square

LEMMA 4.10. *Let H denote a Hankel operator satisfying (A) and choose $(h_m)_{m \in \mathbb{N}} \subseteq L^1(\mathbb{R}^+; B(\mathcal{U}, \mathcal{Y}))$ such that*

$$(4.26) \quad h_m \xrightarrow{L^1} h, \quad \text{as } m \rightarrow \infty.$$

Defining the operators H_m by (4.2), there exist constants $C_1, C_2 > 0$ such that, for $m \in \mathbb{N}$ sufficiently large,

$$(4.27) \quad \|H^*H - H_m^*H_m\|_{1,1} \leq C_1\|h - h_m\|_1,$$

$$(4.28) \quad \|HH^* - H_mH_m^*\|_{1,1} \leq C_2\|h - h_m\|_1.$$

Thus $H_m^* H_m$ and $H_m H_m^*$ converge uniformly to $H^* H$ and $H H^*$, respectively, on $W^{1,1}$ as m tends to infinity.

Proof. By construction H_m satisfy (A) with h replaced by h_m and so the conclusions of Lemma 4.9 hold for H_m, H_m^* and the Schmidt pairs of H_m . We prove the estimate (4.27); the proof of (4.28) is similar and is thus omitted. For $v \in W^{1,1}$ consider

$$(4.29) \quad \begin{aligned} \|(H^* H - H_m^* H_m)v\|_{1,1} &\leq \|H^*\|_{1,1} \cdot \|(H - H_m)v\|_{1,1} \\ &\quad + \|H^* - H_m^*\|_{1,1} \cdot \|H_m v\|_{1,1}. \end{aligned}$$

To bound (4.29) we use (4.23) and its versions for H^*, H_m, H^* and the differences $H - H_m$ and $H^* - H_m^*$, namely,

$$\begin{aligned} \|H^*\|_{1,1} &\leq 3\|h^*\|_1 = 3\|h\|_1, \quad \|(H - H_m)v\|_{1,1} \leq 3\|h - h_m\|_1 \cdot \|v\|_{1,1}, \\ \|H_m v\|_{1,1} &\leq 3\|h_m\|_1 \cdot \|v\|_{1,1}, \quad \|H^* - H_m^*\|_{1,1} \leq 3\|h - h_m\|_1. \end{aligned}$$

Applying these bounds in (4.29) gives

$$\begin{aligned} \|(H^* H - H_m^* H_m)v\|_{1,1} &\leq 3(\|h\|_1 + \|h_m\|_1) \cdot \|h - h_m\|_1 \cdot \|v\|_{1,1}, \\ &\leq 3(\|h\|_1 + 2\|h\|_1) \cdot \|h - h_m\|_1 \cdot \|v\|_{1,1}, \quad m \text{ sufficiently large,} \end{aligned}$$

which proves (4.27). \square

We now have all the ingredients to prove Theorem 4.4.

Proof of Theorem 4.4. (i) Follows immediately from Lemma 4.9 applied to H_m .

(iii) From Lemma 4.9 we have H, H_m compact $L^2(\mathbb{R}^+; \mathcal{U}) \rightarrow L^2(\mathbb{R}^+; \mathcal{Y})$. Applying the estimate (4.18) it follows that

$$\|H - H_m\|_2 \leq \|h - h_m\|_1,$$

and so H_m converges uniformly to H on L^2 as $m \rightarrow \infty$. Therefore, by [26, Lemma 10.21], the distinct singular values $\sigma_k^{(m)}$ of H_m with corresponding multiplicities $p_k^{(m)}$ converge as in (4.3) and moreover we can choose orthonormal Schmidt pairs of H_m that converge in L^2 to (a basis of) orthonormal Schmidt pairs of H as claimed.

(ii) The result of Lemma 4.10 combined with the L^2 convergence established in (iii) implies that all the hypotheses of Lemma 4.8 hold with T, \mathcal{B} , and \mathcal{H} given by (4.25) and $T_m = H_m^* H_m$ (respectively, $T_m = H_m H_m^*$). Therefore we obtain convergence of a subsequence of the Schmidt pairs in $\mathcal{B}_{\mathcal{U}}$ (and $\mathcal{B}_{\mathcal{Y}}$) which gives the desired convergence in $W^{1,1}$. We use an induction and diagonal sequence argument to obtain the existence of a single subsequence along which every Schmidt vector converges. Specifically, using the above argument we find a subsequence $(\tau_1(m))_{m \in \mathbb{N}}$ along which

$$\left. \begin{aligned} v_{i,q}^{(\tau_1(m))} &\xrightarrow{W^{1,1}} v_{1,r}, \\ w_{i,q}^{(\tau_1(m))} &\xrightarrow{W^{1,1}} w_{1,r} \end{aligned} \right\} \quad \forall i \in \{1, 2, \dots, l_1\}, \forall q, r.$$

Using the above argument again we obtain a subsequence of $(\tau_1(m))_{m \in \mathbb{N}}$, denoted $(\tau_2(m))_{m \in \mathbb{N}}$, such that

$$\left. \begin{aligned} v_{i,q}^{(\tau_2(m))} &\xrightarrow{W^{1,1}} v_{2,r}, \\ w_{i,q}^{(\tau_2(m))} &\xrightarrow{W^{1,1}} w_{2,r} \end{aligned} \right\} \quad \forall i \in \{l_1 + 1, 2, \dots, l_2\}, \forall q, r.$$

By repeating this process we obtain a sequence of subsequences indexed by $\tau_n(m)$, and taking the diagonal sequence $(\tau_m(m))_{m \in \mathbb{N}}$ gives the desired result. \square

4.1. Relation to earlier work. We briefly explore the consequences of the extra assumptions in [13] on both H and its Schmidt pairs. The key difference is whether $h \in L^2$ or not. Recall that Example 2.6 contains a physically motivated example where $h \notin L^2$.

LEMMA 4.11. *Let H denote a Hankel operator satisfying (A). The following are equivalent:*

- (1) $h \in L^2(\mathbb{R}^+; B(\mathcal{U}, \mathcal{Y}))$;
- (2) $H : W^{1,2}(\mathbb{R}^+; \mathcal{U}) \rightarrow W^{1,2}(\mathbb{R}^+; \mathcal{Y})$ is compact.

If either (1) or (2) above hold then

- (3) every Schmidt pair (v, w) of H satisfies $v \in W^{1,2}(\mathbb{R}^+; \mathcal{U})$, $w \in W^{1,2}(\mathbb{R}^+; \mathcal{Y})$.

If additionally the vectors $(v_{i,k}(0))_{i \in \mathbb{N}}^{1 \leq k \leq p_i}$ span \mathcal{U} then (3) implies (1).

Remark 4.12. The assumption $(v_{i,k}(0))_{i \in \mathbb{N}}^{1 \leq k \leq p_i}$ span \mathcal{U} is always the case if \mathcal{U} is one dimensional, as for every $i \in \mathbb{N}$ there always exists k such that $v_{i,k}(0) \neq 0$. See [6, Lemma 4.3] for a proof of this assertion when the singular values are simple and [1, Theorem 7.2] for the general case.

Proof of Lemma 4.11. (1) \Rightarrow (2): The proof is similar to the proof that H is compact on $W^{1,1}$ from Lemma 4.9, only now taking L^2 norms instead of L^1 norms. Note that the same formula (4.19) holds for the derivative of Hf when $f \in W^{1,2}$.

(2) \Rightarrow (1): Rearranging the derivative formula (4.19) gives

$$(4.30) \quad h(t)f(0) = -(Hf)'(t) - (H\dot{f})(t) \quad \forall f \in W^{1,2}(\mathbb{R}^+; \mathcal{U}).$$

The right-hand side of (4.30) is in L^2 , and hence so is the left-hand side. Since \mathcal{U} is finite dimensional, it follows that $h \in L^2$.

(1) or (2) \Rightarrow (3): This is analogous to Lemma 4.9 with $W^{1,1}$ replaced by $W^{1,2}$ and follows in the same way from Lemma 4.6.

Now we assume that $(v_{i,k}(0))_{i \in \mathbb{N}}^{1 \leq k \leq p_i}$ span \mathcal{U} .

(3) \Rightarrow (1): That $h \in L^2$ follows readily from the derivative formula (4.30) with $f = v_{i,k}$ as here

$$h(t)v_{i,k}(0) = -(Hv_{i,k})'(t) - (H\dot{v}_{i,k})(t) = -\sigma_i \dot{w}_{i,k}(t) - (H\dot{v}_{i,k})(t).$$

The right-hand side is L^2 and thus so is the left-hand side. Since this holds on a basis for \mathcal{U} we conclude that $h \in L^2$ as required. \square

The significance of the Schmidt vectors belonging to $W^{1,2}$ is that they belong to the domain of the generator of a semigroup of a well-posed realization of H on L^2 . We describe realizations in section 5.

As we might expect, when $h \in L^2$ and is approximated by h_m in both L^1 and L^2 , we also get convergence of the Schmidt pairs in $W^{1,2}$ which is described in the following corollary.

COROLLARY 4.13. *Let H denote a Hankel operator satisfying (A) and suppose additionally that $h \in L^2$. If $(h_m)_{m \in \mathbb{N}}$ are chosen such that*

$$(4.31) \quad h_m \xrightarrow{L^1, L^2} h, \quad \text{as } m \rightarrow \infty,$$

and H_m are given by (4.2) then all the conclusions of Theorem 4.4 hold. Moreover, the choice of Schmidt pairs of H_m in Theorem 4.4 converge to the Schmidt pairs of H in $W^{1,2}$ as well as the senses already established in (4.4).

Proof. The proof is the same as that of Theorem 4.4 only with the Banach spaces $\mathcal{B}_{\mathcal{Y}}$ in (4.25) replaced by

$$(4.32) \quad \tilde{\mathcal{B}}_{\mathcal{Y}} := W^{1,2} \cap W^{1,1}(\mathbb{R}^+; \mathcal{Y}), \quad \|\cdot\|_{\tilde{\mathcal{B}}_{\mathcal{Y}}} := \|\cdot\|_{1,2} + \|\cdot\|_{1,1}, \quad \mathcal{Y} \in \{\mathcal{U}, \mathcal{Y}\}.$$

Here we restrict attention to $\tilde{\mathcal{B}}_{\mathcal{U}}$, H^*H , and $H_m^*H_m$. Arguing as in Lemma 4.9 and using Lemma 4.11, it follows that H^*H and $H_m^*H_m$ are compact on $\tilde{\mathcal{B}}_{\mathcal{U}}$. Furthermore, a calculation shows that there exists constants $C_3, C_4 > 0$ such that

$$\|H^*H - H_m^*H_m\|_{B(\tilde{\mathcal{B}}_{\mathcal{U}})} \leq C_3\|h - h_m\|_1 + C_4\|h - h_m\|_2.$$

Assumption (4.31) implies that $H_m^*H_m$ converges uniformly to H^*H on $\tilde{\mathcal{B}}_{\mathcal{U}}$ as m tends to infinity. The rest of the proof proceeds as before, just noting that convergence in $\tilde{\mathcal{B}}_{\mathcal{U}}$ implies convergence in $W^{1,1}$ and $W^{1,2}$ via (4.32). \square

Remark 4.14. In [13] the authors choose sequences of partial sums of the Coifman and Rochberg decompositions (see Proposition 3.4) as approximations of h and G . This guarantees nuclear convergence of the Hankel operators, and so L^1 convergence of the kernels and H^∞ convergence of the transfer functions (see the inequalities (6.3) for a proof of these assertions). In [13, Lemma 4.2], the authors tweak the approximating sequence G^m by setting

$$F^m(s) := \frac{G^m(s)}{1 + \varepsilon_m s}, \quad m \in \mathbb{N},$$

for some sequence of positive numbers $(\varepsilon_m)_{m \in \mathbb{N}}$ converging to zero. The sequence $(F^m)_{m \in \mathbb{N}}$ converges to G in the above senses, but also in H^2 . Therefore the impulse responses converge in L^1 and L^2 . We remark in section 5.4 how L^2 convergence of the impulse responses is used in [13]. We remark here though that in light of Corollary 4.13, it follows that the Schmidt pairs of the Hankel operators corresponding to F_m converge in $W^{1,2}$ to those of the Hankel operator corresponding to G .

Remark 4.15. In [13] the space of absolutely continuous, uniformly bounded functions with distributional derivatives in L^1 is used and denoted by C^1 with the norm

$$\|f\|_{C^1} := \|f\|_\infty + \|\dot{f}\|_1.$$

The space C^1 is also used by Adamjan, Arov, and Kreĭn in [1]. The estimate (4.1) shows that $W^{1,1}$ is continuously embedded into C^1 . As such we recover from Lemma 4.9 that the Schmidt pairs of a Hankel operator satisfying (A) also belong to C^1 . Additionally, under the assumptions of Theorem 4.4, from that result we see that the Schmidt pairs of H_m converge in C^1 to Schmidt pairs of H . We have chosen to use $W^{1,1}$ instead of C^1 as $W^{1,1}$ is the domain of the main operator of the exactly observable shift realization on L^1 , the realization we use to define truncated systems in terms of Schmidt pairs.

5. Realizations and truncated realizations of integral Hankel operators.

Here we describe realizations of bounded Hankel operators of the integral form (2.1) with L^1 kernel. We also describe truncations of these realizations, and make precise what we mean by the reduced order system obtained by balanced truncation. Our ultimate aim is to prove Theorem 2.4 and also to provide the ingredients to prove Theorem 2.3. We prove Theorem 2.4 by finding a realization of G_n^m that converges to a realization of G_n . We have a similar strategy to that of [6] and [13], in that we seek a realization that we can describe in terms of the Schmidt pairs of the Hankel operator. Our novel approach is then to use the $W^{1,1}$ convergence of the Schmidt pairs established in section 4. Propositions 5.10 and 5.12 are the main results of this section; the former is a more detailed version of Theorem 2.4 and describes convergence properties of approximate balanced truncations to the exact balanced truncation. The latter describes properties of the reduced order system.

We remind the reader that a version of Theorem 2.4 has been proven for a specific approximation in [13] under the stronger assumptions listed in section 2.1. We are only assuming that the Hankel operator satisfies (A) from section 4 and seek to derive convergence in H^∞ of *any* approximate sequence of reduced order transfer functions G_n^m , satisfying $h_m \rightarrow h$ in L^1 as $m \rightarrow \infty$, to the exact reduced order transfer function G_n .

We briefly recap well-posed linear systems and realizations. Well-posed linear systems on L^2 date back to the work of Salamon [34, 35] and Weiss [38, 39]. We use the notation

$$(5.1) \quad \left[\begin{array}{c|c} \mathbb{T} & \Phi \\ \hline \Psi & \mathbb{F} \end{array} \right] \quad \text{on } (\mathcal{Y}, \mathcal{X}, \mathcal{U})$$

to denote an L^p well-posed linear system for $1 \leq p < \infty$ with output space, state space, and input space \mathcal{Y} , \mathcal{X} , and \mathcal{U} , respectively. Here \mathbb{T} is a strongly continuous semigroup, Φ is the input operator, Ψ is the output operator, and \mathbb{F} is the input-output operator. For example, if $\begin{bmatrix} A & B \\ C & D \end{bmatrix}$ are the generators of a finite-dimensional input-state-output system then

$$\left. \begin{aligned} \mathbb{T}(t) &= e^{At}, & (\Phi u)(t) &= \int_0^t e^{A(t-s)} Bu(s) ds, \\ (\Psi x_0)(t) &= Ce^{At}x_0, & (\mathbb{F}u)(t) &= Du(t) + C \int_0^t e^{A(t-s)} Bu(s) ds, \end{aligned} \right\} t \geq 0.$$

We will make use of many results from the monograph of Staffans [37] dedicated to well-posed linear systems. We refer the reader to [37, section 2.8] for the precise definition of (5.1) and how our notation is related to the notation of [37].

The term realization is usually understood as a realization of an input-output (linear, time-invariant, causal) map on L^p , rather than of a Hankel operator. We refer the reader to [37, Definition 2.6.3] for more details. We remark, however, that a system with impulse response $h \in L^1(\mathbb{R}^+; B(\mathcal{U}; \mathcal{Y}))$ has a transfer function which is regular in the uniform topology with zero feedthrough by [37, Theorem 5.6.7]. Recall from section 3 that the transfer function is only determined by the Hankel operator up to an arbitrary constant, the feedthrough. By ensuring $h \in L^1$ we have fixed zero feedthrough and so the Hankel operator completely determines the transfer function. Therefore for the class of systems we consider, a realization of the transfer function is equivalent to a realization of the Hankel operator.

Balanced realizations of a system with a Hankel operator satisfying the assumptions of section 2.1 are described in [6, section 2]. Output-normal realizations for the same class of systems are described in [13, section 3]. More recently, output-normal and balanced realizations have been described for L^2 well-posed linear systems in Staffans [37, Chapter 9].

We now describe the realization we use that is similar in effect to an output-normal realization, but instead has a Banach space state space. We are unable to use the realizations in [6] or [13] as the impulse response is not necessarily in L^2 . We describe in more detail in section 5.4 why this is important.

5.1. The exactly observable shift realization on L^1 . From [37, Example 2.6.5 (ii)] any bounded Hankel operator $H : L^1(\mathbb{R}^+; \mathcal{U}) \rightarrow L^1(\mathbb{R}^+; \mathcal{Y})$ given by (2.1) has an L^1 well-posed shift realization

$$(5.2) \quad \left[\begin{array}{c|c} \mathbb{S}^* & H \\ \hline I & \mathbb{F} \end{array} \right] \quad \text{on } (\mathcal{Y}, L^1(\mathbb{R}^+; \mathcal{U}), \mathcal{U}),$$

called the exactly observable shift realization on L^1 , where \mathbb{F} is the input-output map, which by the above discussion is determined entirely by the Hankel operator. The semigroup \mathbb{S}^* is the left-shift semigroup on $L^1(\mathbb{R}^+; \mathcal{Y})$, defined analogously to that in Definition 3.2. The generating operators of the realization (5.2) are given by the next lemma.

LEMMA 5.1. *Let H denote a Hankel operator satisfying (A). The generating operators (A, B, C) of the exactly observable shift realization (5.2) of H are*

$$(5.3) \quad A : D(A) \rightarrow L^1(\mathbb{R}^+; \mathcal{Y}), \quad Aw = \dot{w}, \quad w \in D(A) = W^{1,1}(\mathbb{R}^+; \mathcal{Y}),$$

$$(5.4) \quad B : \mathcal{U} \rightarrow L^1(\mathbb{R}^+; \mathcal{Y}), \quad Bu = h(\cdot)u, \quad u \in \mathcal{U},$$

$$(5.5) \quad C : D(A) \rightarrow \mathcal{Y}, \quad Cx = x(0), \quad x \in D(A).$$

Proof. By [37, Example 3.2.3 (ii)] the operator A given in (5.3) is the generator of the left shift semigroup \mathbb{S}^* . Similarly, by [37, Example 4.4.6] the operator C in (5.5) is the observation operator of (5.2). To find the control operator first note that B defined in (5.4) is certainly bounded as $h \in L^1$. The (extended) input map of the system with generators (A, B) is (formally) given by

$$L^1(\mathbb{R}^+; \mathcal{U}) \ni u \mapsto \Phi_\infty u = \int_{\mathbb{R}^+} \mathbb{S}^*(s)Bu(s) ds.$$

Using the formula (5.4) we see that for $u \in L^1$ and $t \geq 0$

$$(\Phi_\infty u)(t) = \left[\int_{\mathbb{R}^+} \mathbb{S}^*(s)Bu(s) ds \right] (t) = \int_{\mathbb{R}^+} h(t+s)u(s) ds = (Hu)(t),$$

that is, Φ_∞ is well-defined as it is equal to the Hankel operator H , which is the input map for the realization (5.2). Since the control operator in $B(\mathcal{U}, L^1(\mathbb{R}^+; \mathcal{Y}))$ is unique, it follows that B defined by (5.4) must be the control operator for the realization (5.2). \square

Remark 5.2. The exactly observable shift realization (5.2) is generally not approximately controllable, and so not minimal. However, by [37, Theorem 9.1.9 (i)] we can obtain a minimal realization from (5.2) by changing (reducing) the state space to $\overline{\text{im } H}$, the reachable subspace, instead. That (5.2) is not necessarily controllable is not an issue, as we shall see in section 5.2 that the truncation method gives rise to a minimal finite-dimensional system.

We need the following ‘‘adjoint’’ operators to those of Lemma 5.1.

LEMMA 5.3. *Let H denote a Hankel operator satisfying (A) and let A, B denote the generating operators from Lemma 5.1. The operators defined by*

$$(5.6) \quad \begin{aligned} A^* : D(A^*) \rightarrow L^1(\mathbb{R}^+; \mathcal{Y}), \quad A^*w = -\dot{w}, \quad w \in D(A^*) = W_0^{1,1}(\mathbb{R}^+; \mathcal{Y}), \\ B^* : D(A) \rightarrow \mathcal{U}, \quad B^*x = (H^*x)(0), \quad x \in D(A), \end{aligned}$$

are adjoint to A and B in the sense that

$$(5.7) \quad \begin{aligned} \langle Ax, y \rangle_{L^2} &= \langle x, A^*y \rangle_{L^2} \quad \forall x \in D(A), \forall y \in D(A^*), \\ \langle Bu, x \rangle_{L^2} &= \langle u, B^*x \rangle_{\mathcal{U}} \quad \forall u \in \mathcal{U}, \forall x \in D(A). \end{aligned}$$

The above L^2 inner products are understood as the duality pairing of L^1 and L^∞ (the latter containing $W^{1,1}$). Recall here that $D(A) = W^{1,1}(\mathbb{R}^+; \mathcal{Y})$.

Proof. For the adjoint property (5.7) between A and A^* the key calculation is

$$\langle Ax, y \rangle_{L^2} = \langle \dot{x}, y \rangle_{L^2} = [\langle x(t), y(t) \rangle_{\mathcal{Y}}]_0^\infty - \langle x, \dot{y} \rangle_{L^2},$$

where we have integrated by parts. Now using that $x \in W^{1,1}$,

$$\langle Ax, y \rangle_{L^2} = -\langle x(0), y(0) \rangle_{\mathcal{Y}} - \langle x, \dot{y} \rangle_{L^2} = -\langle x, \dot{y} \rangle_{L^2} = \langle x, A^*y \rangle_{L^2}$$

when $y \in D(A^*) = W_0^{1,1}(\mathbb{R}^+; \mathcal{Y})$. We now consider B^* , which is certainly well-defined on its domain as for $x \in D(A)$

$$\|B^*x\|_{\mathcal{U}} = \|(H^*x)(0)\|_{\mathcal{U}} \leq \|h^*\|_1 \cdot \|x\|_\infty \leq \|h\|_1 \cdot \|x\|_{1,1}.$$

To see the adjoint property (5.7)

$$\begin{aligned} \langle Bu, x \rangle_{L^2} &= \int_{\mathbb{R}^+} \langle h(s)u, x(s) \rangle_{\mathcal{Y}} ds = \left\langle u, \int_{\mathbb{R}^+} h^*(s)x(s) ds \right\rangle_{\mathcal{U}} \\ &= \langle u, (H^*x)(0) \rangle_{\mathcal{U}} = \langle u, B^*x \rangle_{\mathcal{U}}. \quad \square \end{aligned}$$

5.2. Truncations of the exactly observable shift realization. The following lemma and definition forms a more detailed version of Definition 2.2.

LEMMA 5.4. Let $(w_{i,k})_{i \in \mathbb{N}}^{1 \leq k \leq p_i}$ denote an orthonormal basis of Schmidt vectors of a Hankel operator satisfying (A). For $n \in \mathbb{N}$ define

$$(5.8) \quad \mathcal{X}_n := \text{span} \{w_{i,k} \mid 1 \leq i \leq n, 1 \leq k \leq p_i\},$$

which is a closed subspace of $L^1, W^{1,1}$, and L^2 . We use the notation $\mathcal{X}_n^1, \mathcal{X}_n^{1,1}$, and \mathcal{X}_n^2 to denote \mathcal{X}_n considered as a subspace of $L^1, W^{1,1}$, and L^2 , respectively. Then there exist complementary subspaces $\mathcal{X}_n^1, \mathcal{X}_n^{1,1}$, and \mathcal{X}_n^2 such that

$$\begin{aligned} L^1(\mathbb{R}^+; \mathcal{Y}) &= \mathcal{X}_n^1 \oplus \mathcal{X}_n^1, \\ W^{1,1}(\mathbb{R}^+; \mathcal{Y}) &= \mathcal{X}_n^{1,1} \oplus \mathcal{X}_n^{1,1}, \\ L^2(\mathbb{R}^+; \mathcal{Y}) &= \mathcal{X}_n^2 \oplus \mathcal{X}_n^2, \end{aligned}$$

and these decompositions are all orthogonal with respect to the L^2 inner product or duality product as appropriate. There exist continuous projections

$$\begin{aligned} \mathbb{P}_n : L^2(\mathbb{R}^+; \mathcal{Y}) &\rightarrow \mathcal{X}_n^2, & \mathbb{Q}_n &:= I - \mathbb{P}_n : L^2(\mathbb{R}^+) \rightarrow \mathcal{X}_n^2, \\ \mathbb{P}_n : W^{1,1}(\mathbb{R}^+; \mathcal{Y}) &\rightarrow \mathcal{X}_n^{1,1}, & \mathbb{Q}_n &:= I - \mathbb{P}_n : W^{1,1}(\mathbb{R}^+) \rightarrow \mathcal{X}_n^{1,1}, \\ \mathbb{P}_n : L^1(\mathbb{R}^+; \mathcal{Y}) &\rightarrow \mathcal{X}_n^1, & \mathbb{Q}_n &:= I - \mathbb{P}_n : L^1(\mathbb{R}^+) \rightarrow \mathcal{X}_n^1, \end{aligned}$$

where \mathbb{P}_n is a restriction of \mathbb{P}_n and \mathbb{P}_n is the continuous extension of \mathbb{P}_n . Each projection $\mathbb{P}_n, \mathbb{P}_n$, and \mathbb{P}_n is given by

$$(5.9) \quad x \mapsto \sum_{i=1}^n \sum_{k=1}^{p_i} \langle w_{i,k}, x \rangle_{L^2} w_{i,k}$$

on its domain. The projections $\mathbb{P}_n, \mathbb{Q}_n, \mathbb{P}_n$ and \mathbb{Q}_n satisfy the self-adjoint-like relations

$$(5.10) \quad \left. \begin{aligned} \langle x, \mathbb{P}_n y \rangle_{L^2} &= \langle \mathbb{P}_n x, y \rangle_{L^2}, \\ \langle x, \mathbb{Q}_n y \rangle_{L^2} &= \langle \mathbb{Q}_n x, y \rangle_{L^2} \end{aligned} \right\} \quad \forall x, y \in L^1(\mathbb{R}^+; \mathcal{Y}),$$

$$(5.11) \quad \left. \begin{aligned} \langle x, \mathbb{P}_n y \rangle_{L^2} &= \langle \mathbb{P}_n x, y \rangle_{L^2}, \\ \langle x, \mathbb{Q}_n y \rangle_{L^2} &= \langle \mathbb{Q}_n x, y \rangle_{L^2} \end{aligned} \right\} \quad \forall x, y \in W^{1,1}(\mathbb{R}^+; \mathcal{Y}).$$

The L^2 inner products in (5.10) are understood as the duality pairing.

Proof. We do not give the full proof as it is reasonably elementary. Interested readers should consult Guiver [17, Appendix B]. The decomposition of L^2 follow from the usual orthogonal decomposition. Since $\mathcal{X}_n \subseteq W^{1,1} \subseteq L^2$ we obtain the decomposition of $W^{1,1}$ by restriction. As $W^{1,1} \subseteq L^1$ we extend the decomposition of $W^{1,1}$ by continuity to obtain that for L^1 . Briefly, the self-adjointness properties follow from the orthogonality of the decompositions with respect to the L^2 inner product. \square

We now have the ingredients to define what we mean by a balanced truncation.

DEFINITION 5.5. *Let (A, B, C) denote the generating operators from Lemma 5.1 of the realization (5.2) of a Hankel operator satisfying assumption (A). Using the decompositions and projections of Lemma 5.4 we define the operators*

$$(5.12) \quad A_n := \mathcal{P}_n A|_{\mathcal{X}_n^{1,1}}, \quad B_n := \mathcal{P}_n B, \quad C_n := C|_{\mathcal{X}_n^{1,1}}.$$

The operators in (5.12) generate a finite-dimensional linear system on $(\mathcal{Y}, \mathcal{X}_n, \mathcal{U})$, called the reduced order system obtained by balanced truncation, or just the balanced truncation, which we denote by $\begin{bmatrix} A_n & B_n \\ C_n & 0 \end{bmatrix}$. The function

$$(5.13) \quad s \mapsto G_n(s) := C_n(sI - A_n)^{-1}B_n,$$

defined and analytic on some right-half complex plane, is called the reduced order transfer function obtained by balanced truncation.

Remark 5.6.

- (i) For the operators defined in (5.12) to make sense it is crucial that $\mathcal{X}_n \subseteq D(A)$ and that B is bounded; properties established in Lemmas 4.9 and 5.1, respectively.
- (ii) If G is irrational then H has infinitely many nonzero (and thus positive) singular values. The sequence of distinct singular values $(\sigma_k)_{k \in \mathbb{N}}$ may or may not be infinite, as a given singular value may have infinite multiplicity. When additionally H is nuclear, however, $(\sigma_k)_{k \in \mathbb{N}}$ has infinitely many terms and these values must converge to zero. In this case the truncation space \mathcal{X}_n in (5.8) and the balanced truncation $\begin{bmatrix} A_n & B_n \\ C_n & 0 \end{bmatrix}$ from Definition 5.5 are defined for every $n \in \mathbb{N}$. In this work our standing assumptions for balanced truncation is that G is irrational and H is nuclear. To prove our results, however, we shall use rational functions and their balanced truncations as intermediate approximations. A transfer function J is rational precisely when its Hankel operator H_J is finite rank and therefore its sequence of singular values contains only finitely many nonzero terms. In this case the truncation space and balanced truncation are only defined for $n \leq \mathcal{N}$, where \mathcal{N} is the number of distinct singular values of H_J .
- (iii) In Lemma 5.4 we define \mathcal{X}_n as the direct sum of eigenspaces of H^*H corresponding to the first n eigenvalues, which are assumed from section 2, and throughout this work, the n largest eigenvalues. Recall that the square roots of the eigenvalues of H^*H are the singular values of H . Keeping the largest singular values in the truncated system, and omitting the rest, is essential for a tighter error bound in (2.4). In principle, however, we could define a truncated system as in Definition 5.5 by restricting and projecting onto *any* sum of eigenspaces.

Remark 5.7. We now drop the distinction $\mathcal{X}_n^1, \mathcal{X}_n^{1,1}, \mathcal{X}_n^2$ and simply consider A_n as an operator $A_n : \mathcal{X}_n \rightarrow \mathcal{X}_n$, where \mathcal{X}_n is still given by (5.8) and is equipped with the L^2 inner product, so that $(\mathcal{X}_n, \|\cdot\|_2)$ is a finite-dimensional Hilbert space.

Remark 5.8. Let $\mathbf{u} := \dim \mathcal{U}$ and $\mathbf{y} := \dim \mathcal{Y}$ and choose orthonormal bases $(y_i)_{i=1}^{\mathbf{y}}$, $(w_{i,k})_{\substack{1 \leq k \leq p_i \\ 1 \leq i \leq n}}$, and $(u_i)_{i=1}^{\mathbf{u}}$ for \mathcal{Y} , \mathcal{X}_n , and \mathcal{U} , respectively. Then the operators (A_n, B_n, C_n) in (5.12) have (block) matrix representations with respect to the above bases:

$$(5.14) \quad \begin{aligned} \mathcal{A}(n) &:= (\mathcal{A}_{ij})_{i,j=1}^n, & \mathcal{A}_{ij} &\in \mathbb{C}^{p_i \times p_j}, & (\mathcal{A}_{ij})_{kl} &= \langle w_{i,k}, w_{j,l} \rangle_{L^2}, \\ \mathcal{B}(n) &:= (\mathcal{B}_i)_{i=1}^n, & \mathcal{B}_i &\in \mathbb{C}^{p_i \times \mathbf{u}}, & (\mathcal{B}_i)_{kl} &= \langle \sigma_i v_{i,k}(0), u_l \rangle_{\mathcal{U}}, \\ \mathcal{C}(n) &:= (\mathcal{C}_i)_{i=1}^n, & \mathcal{C}_i &\in \mathbb{C}^{\mathbf{y} \times p_i}, & (\mathcal{C}_i)_{kl} &= \langle y_k, w_{i,l}(0) \rangle_{\mathcal{Y}}. \end{aligned}$$

We recall the definition of a stable and output-normal realization for finite-dimensional systems. The latter concept extends to infinite-dimensional systems, but we do not need it for our purposes.

DEFINITION 5.9. Let $\begin{bmatrix} A & B \\ C & D \end{bmatrix}$ denote a minimal realization of a rational transfer function. We say that

- (i) A or $\begin{bmatrix} A & B \\ C & D \end{bmatrix}$ is stable if A is asymptotically stable, that is, every eigenvalue of A has negative real part, and
- (ii) $\begin{bmatrix} A & B \\ C & D \end{bmatrix}$ is output normal if the observability Gramian is the identity.

We now have the ingredients to state and prove the first of our two main results of this section. Both of these results are used in the proof of the error bound of Theorem 2.3. Proposition 5.10 below is a more detailed version of Theorem 2.4 and describes convergence properties of approximate balanced truncations to the exact balanced truncation.

PROPOSITION 5.10. Let H denote a Hankel operator satisfying assumption (A) with transfer function G . Choose orthonormal bases $(y_i)_{i=1}^{\mathbf{y}}$ and $(u_i)_{i=1}^{\mathbf{u}}$ for \mathcal{Y} and \mathcal{U} , respectively, where $\mathbf{y} = \dim \mathcal{Y}$ and $\mathbf{u} = \dim \mathcal{U}$. Let $(h_m)_{m \in \mathbb{N}}$ denote any sequence in $L^1(\mathbb{R}^+; B(\mathcal{U}; \mathcal{Y}))$, chosen such that $h_m \rightarrow h$ in L^1 as $m \rightarrow \infty$. Let $(G^m)_{m \in \mathbb{N}}$ denote the sequence of transfer functions each with impulse response h_m and let $\mathcal{N}(m)$ denote the number of distinct singular values of the Hankel operator of G^m . Let (A^m, B^m, C^m) denote the generators from Lemma 5.1 of the exactly observable shift realization (5.2) of G^m . For $n \in \mathbb{N}$ and $m \in \mathbb{N}$ with $n \leq \mathcal{N}(m)$, let (A_n^m, B_n^m, C_n^m) denote the balanced truncation of G^m from Definition 5.5 on $(\mathcal{Y}, \mathcal{X}_n^m, \mathcal{U})$, which is well-defined by Remark 5.6(ii). If the Schmidt vectors defining \mathcal{X}_n^m are chosen as in Theorem 4.4 then there exists a subsequence $(\tau(m))_{m \in \mathbb{N}}$, such that the following hold:

- (1) the matrix representations of $A_n^{\tau(m)}$, $B_n^{\tau(m)}$, and $C_n^{\tau(m)}$ with respect to the bases

$$(y_i)_{i=1}^{\mathbf{y}}, \quad (w_{i,k}^{(\tau(m))})_{\substack{1 \leq k \leq p_i^{(\tau(m))} \\ 1 \leq i \leq l_n}}, \quad \text{and} \quad (u_i)_{i=1}^{\mathbf{u}} \quad \text{for } \mathcal{Y}, \mathcal{X}_n^{\tau(m)}, \text{ and } \mathcal{U}$$

converge elementwise to matrix representations of A_n, B_n , and C_n with respect to the bases

$$(y_i)_{i=1}^{\mathbf{y}}, \quad (w_{i,k})_{\substack{1 \leq k \leq p_i \\ 1 \leq i \leq n}}, \quad \text{and} \quad (u_i)_{i=1}^{\mathbf{u}} \quad \text{for } \mathcal{Y}, \mathcal{X}_n, \text{ and } \mathcal{U}.$$

The operators A_n, B_n , and C_n are truncated operators from Definition 5.5 of the exactly observable shift realization (5.2) for H ;

- (2) letting $G_n^{\tau(m)}$ and G_n denote the reduced order transfer functions obtained by balanced truncation from $G^{\tau(m)}$ and G , respectively, then

$$(5.15) \quad G_n^{\tau(m)} \xrightarrow{H^\infty} G_n, \quad \text{as } m \rightarrow \infty.$$

Remark 5.11.

- (i) Under the assumptions of Proposition 5.10, if additionally the singular values of H are simple, then all the convergence in Proposition 5.10 holds without needing a subsequence.
- (ii) For ease of presentation, we only prove the first claim of Proposition 5.10 in the case when the singular values of H and H_m (the Hankel operator with kernel h_m) are simple. A proof of the general case can be found in [17, Proposition 5.3.9]. In either case our strategy is the same: to prove the convergence in (1) by proving componentwise convergence of the matrix representations of A_n^m, B_n^m , and C_n^m from Remark 5.8. The notation in the general case, however, becomes so cumbersome as to obscure the argument.

Proof of Proposition 5.10. From Theorem 4.4 and Remark 4.5(ii) every Schmidt vector of H_m converges in L^2 and $W^{1,1}$ to a Schmidt vector of H . Thus for every $k \in \mathbb{N}$,

$$(5.16) \quad \begin{aligned} v_k^{(m)} &\xrightarrow{W^{1,1}} v_k, \\ w_k^{(m)} &\xrightarrow{W^{1,1}} w_k, \end{aligned} \quad \text{as } m \rightarrow \infty.$$

Furthermore, from Theorem 4.4 for $n \in \mathbb{N}$ and $1 \leq k \leq n$, the w_k form an orthonormal (in L^2) basis for \mathcal{X}_n given by (5.8). Define matrices $(\mathcal{A}(n), \mathcal{B}(n), \mathcal{C}(n))$ by (5.14), with entries in terms of the above $W^{1,1}$ limits. Since these $W^{1,1}$ limits are Schmidt pairs of H , it follows that $(\mathcal{A}(n), \mathcal{B}(n), \mathcal{C}(n))$ are the matrix representations (with respect to the bases $(y_i)_{i=1}^y, (w_i)_{i=1}^n$, and $(u_i)_{i=1}^u$) of A_n, B_n , and C_n , respectively.

Under the assumption that the singular values are simple, the formulas (5.14) simplify to

$$(5.17) \quad \begin{aligned} \mathcal{A}(n) &\in \mathbb{C}^{n \times n}, & (\mathcal{A}(n))_{ij} &= \langle w_i, \dot{w}_j \rangle_{L^2}, \\ \mathcal{B}(n) &\in \mathbb{C}^{n \times u}, & (\mathcal{B}(n))_{ij} &= \langle \sigma_i v_i(0), u_j \rangle_{\mathcal{U}}, \\ \mathcal{C}(n) &\in \mathbb{C}^{y \times n}, & (\mathcal{C}(n))_{ij} &= \langle y_i, w_j(0) \rangle_{\mathcal{Y}}. \end{aligned}$$

The truncations (A_n^m, B_n^m, C_n^m) of (A^m, B^m, C^m) from Definition 5.5 have matrix representations $(\mathcal{A}^m(n), \mathcal{B}^m(n), \mathcal{C}^m(n))$, as in (5.14). Again these simplify to

$$(5.18) \quad \begin{aligned} \mathcal{A}^m(n) &\in \mathbb{C}^{n \times n}, & (\mathcal{A}^m(n))_{ij} &= \langle w_i^{(m)}, \dot{w}_j^{(m)} \rangle_{L^2}, \\ \mathcal{B}^m(n) &\in \mathbb{C}^{n \times u}, & (\mathcal{B}^m(n))_{ij} &= \langle \sigma_i^{(m)} v_i^{(m)}(0), u_j \rangle_{\mathcal{U}}, \\ \mathcal{C}^m(n) &\in \mathbb{C}^{y \times n}, & (\mathcal{C}^m(n))_{ij} &= \langle y_i, w_j^{(m)}(0) \rangle_{\mathcal{Y}}. \end{aligned}$$

We prove that the matrices in (5.18) converge elementwise to (5.17) which proves the first claim. We have for $1 \leq i, j \leq n$

$$\begin{aligned} |(\mathcal{A}(n))_{ij} - (\mathcal{A}^m(n))_{ij}| &= |\langle w_i, \dot{w}_j \rangle_{L^2} - \langle w_i^{(m)}, \dot{w}_j^{(m)} \rangle_{L^2}| \\ &\leq \|w_i - w_i^{(m)}\|_{\infty} \cdot \|\dot{w}_j\|_1 + \|w_i^{(m)}\|_{\infty} \cdot \|\dot{w}_j - \dot{w}_j^{(m)}\|_1 \end{aligned}$$

by the Hölder inequality. Thus by (5.16)

$$\begin{aligned} |(\mathcal{A}(n))_{ij} - (\mathcal{A}^m(n))_{ij}| &\leq \|w_i - w_i^{(m)}\|_{1,1} \cdot \|w_j\|_{1,1} + \|w_i^{(m)}\|_{1,1} \cdot \|w_j - w_j^{(m)}\|_{1,1} \\ &\rightarrow 0, \quad \text{as } m \rightarrow \infty, \text{ by (5.16).} \end{aligned}$$

Next consider for $1 \leq i \leq n, 1 \leq j \leq u$,

$$\begin{aligned} |(\mathcal{B}(n))_{ij} - (\mathcal{B}^m(n))_{ij}| &= \left| \langle \sigma_i v_i(0), u_j \rangle_{\mathcal{U}} - \langle \sigma_i^{(m)} v_i^{(m)}(0), u_j \rangle_{\mathcal{U}} \right| \\ &\leq |\sigma_i^{(m)}| \cdot \|v_i^{(m)}(0) - v_i(0)\|_{\mathcal{U}} \cdot \|u_j\|_{\mathcal{U}} \\ &\quad + |\sigma_i^{(m)} - \sigma_i| \cdot \|v_i(0)\|_{\mathcal{U}} \cdot \|u_j\|_{\mathcal{U}} \end{aligned}$$

by the Cauchy–Schwarz inequality. Hence for m sufficiently large

$$\begin{aligned} |(\mathcal{B}(n))_{ij} - (\mathcal{B}^m(n))_{ij}| &\leq 2|\sigma_i| \cdot \|v_i^{(m)} - v_i\|_{1,1} + |\sigma_i^{(m)} - \sigma_i| \cdot \|v_i\|_{1,1} \\ &\rightarrow 0, \quad \text{as } m \rightarrow \infty. \end{aligned}$$

The convergence of $(\mathcal{C}^m(n))_{ij}$ to $(\mathcal{C}(n))_{ij}$ is proved similarly.

The second claim, the convergence in (5.15), follows once Proposition 5.10(1) is established, as in the proof of [13, Lemma 4.4]. \square

5.3. Properties of the balanced truncation and Lyapunov equations.

We now state and prove the second main result of section 5, which describes properties of the balanced truncation.

PROPOSITION 5.12. *Let H denote a Hankel operator satisfying assumption (A) with transfer function G and let G_n denote the transfer function obtained by balanced truncation of G . The realization $\begin{bmatrix} A_n & B_n \\ C_n & 0 \end{bmatrix}$ on $(\mathcal{Y}, \mathcal{X}_n, \mathcal{U})$ of G_n from Definition 5.5 is minimal, stable, and output normal. Moreover, the Hankel singular values of H , with the same multiplicities.*

The proof of Proposition 5.12 is conceptually very similar to that of Pernebo and Silverman [31] for Lyapunov balanced truncation for finite-dimensional systems. A proof in the finite-dimensional case can also be found in Green and Limebeer [16, Lemma 9.4.1]. The broad idea is to derive some Lyapunov equations that the truncated operators A_n, B_n, C_n and their adjoints satisfy. From here we prove A_n is stable and then the claims that $\begin{bmatrix} A_n & B_n \\ C_n & 0 \end{bmatrix}$ is minimal and output normal follow from standard finite-dimensional arguments. Since the operators to be truncated A, B , and C are defined on Banach spaces with some inherited Hilbert space structure, we argue carefully and need to collect some technical results beforehand.

We first make a remark on the notation we shall use from now on. Recall also the interpretation of \mathcal{X}_n from Remark 5.7 as a Hilbert space equipped with the L^2 inner product.

Remark 5.13. Given a Hankel operator satisfying (A) let A, B, C denote the operators from Lemma 5.1, and recall the decompositions and projections of Lemma 5.4. We define the decompositions

$$(5.19) \quad A = \begin{bmatrix} \mathcal{P}_n A|_{\mathcal{X}_n} & \mathcal{P}_n A|_{\mathcal{X}_n^{1,1}} \\ \mathcal{Q}_n A|_{\mathcal{X}_n} & \mathcal{Q}_n A|_{\mathcal{X}_n^{1,1}} \end{bmatrix} =: \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix},$$

$$(5.20) \quad B = \begin{bmatrix} \mathcal{P}_n B \\ \mathcal{Q}_n B \end{bmatrix} =: \begin{bmatrix} B_1 \\ B_2 \end{bmatrix},$$

$$(5.21) \quad C = [C|_{\mathcal{X}_n} \quad C|_{\mathcal{X}_n^{1,1}}] =: [C_1 \quad C_2],$$

so that $A_n = A_{11}, B_n = B_1$, and $C_n = C_1$.

LEMMA 5.14. *Given the operators and decompositions of Lemmas 5.1 and 5.4 and the notation of Remark 5.13, let ${}^2A_{11}^* : \mathcal{X}_n \rightarrow \mathcal{X}_n$ denote the (Hilbert space)*

adjoint of A_{11} so that

$$(5.22) \quad \langle x, A_{11}y \rangle_{L^2} = \langle {}^2A_{11}^*x, y \rangle_{L^2} \quad \forall x, y \in \mathcal{X}_n.$$

The operator ${}^2A_{11}^*$ is an extension of

$${}^1A_{11}^* := \mathcal{P}_n A^*|_{\mathcal{X}_n \cap D(A^*)} : \mathcal{X}_n \cap D(A^*) \rightarrow \mathcal{X}_n,$$

where A^* is the (adjoint) operator from Lemma 5.3. Therefore ${}^1A_{11}^* \subseteq {}^2A_{11}^*$, which are equal on $\mathcal{X}_n \cap D(A^*)$, and for simplicity we denote both of these operators by A_{11}^* on $\mathcal{X}_n \cap D(A^*)$. Defining the restrictions

$$B_1^* := B^*|_{\mathcal{X}_n} : \mathcal{X}_n \rightarrow \mathcal{U}, \quad B_2^* := B^*|_{\mathcal{X}_n^{1,1}} : \mathcal{X}_n^{1,1} \rightarrow \mathcal{U},$$

the Hilbert space adjoint of B_1^* is $B_1 = \mathcal{P}_n B : \mathcal{U} \rightarrow \mathcal{X}_n$, so that

$$(5.23) \quad \langle x, B_1u \rangle_{L^2} = \langle B_1^*x, u \rangle_{\mathcal{U}} \quad \forall u \in \mathcal{U}, \forall x \in \mathcal{X}_n.$$

Proof. For $x, y \in \mathcal{X}_n$

$$(5.24) \quad \begin{aligned} \langle x, Ay \rangle_{L^2} &= \langle \mathcal{P}_n x, A|_{\mathcal{X}_n} y \rangle_{L^2} = \langle x, \mathcal{P}_n A|_{\mathcal{X}_n} y \rangle_{L^2} \quad \text{by (5.10),} \\ &= \langle x, A_{11}y \rangle_{L^2}. \end{aligned}$$

If additionally $x \in D(A^*)$ then by the adjoint property (5.7)

$$(5.25) \quad \begin{aligned} \langle x, Ay \rangle_{L^2} &= \langle A^*x, y \rangle_{L^2} = \langle A^*|_{\mathcal{X}_n \cap D(A^*)} x, \mathcal{P}_n y \rangle_{L^2}, \\ &= \langle \mathcal{P}_n A^*|_{\mathcal{X}_n \cap D(A^*)} x, y \rangle_{L^2} \quad \text{by (5.10),} \\ &= \langle {}^1A_{11}^*x, y \rangle_{L^2}. \end{aligned}$$

Comparing (5.24) and (5.25) we obtain

$$(5.26) \quad \langle x, A_{11}y \rangle_{L^2} = \langle {}^1A_{11}^*x, y \rangle_{L^2} \quad \forall x \in \mathcal{X}_n \cap D(A^*), \quad \forall y \in \mathcal{X}_n.$$

The Hilbert space adjoint ${}^2A_{11}^*$ satisfies (5.22) by definition, and so the claims of the lemma follow from (5.26) and the unicity of the Hilbert space adjoint.

To prove the claims for B_1^* it suffices to prove (5.23). Let $u \in \mathcal{U}$, $x \in \mathcal{X}_n$ so that

$$\begin{aligned} \langle B_1^*x, u \rangle_{\mathcal{U}} &= \langle B^*x, u \rangle_{\mathcal{U}} = \langle x, Bu \rangle_{\mathcal{U}} \quad \text{by (5.7),} \\ &= \langle \mathcal{P}_n x, Bu \rangle_{\mathcal{U}} = \langle x, \mathcal{P}_n Bu \rangle_{\mathcal{U}} \quad \text{by (5.10),} \\ &= \langle x, B_1u \rangle_{\mathcal{U}}, \end{aligned}$$

and so the result follows by the unicity of the Hilbert space adjoint of B_1^* . \square

DEFINITION 5.15. Given the operators of Lemma 5.1 and decompositions of Lemma 5.4, recall the operator A_{12} from Remark 5.13, given by

$$A_{12} = \mathcal{P}_n A|_{\mathcal{X}_n^{1,1}} : \mathcal{X}_n^{1,1} \rightarrow \mathcal{X}_n.$$

We denote by A_{12}^* the operator $\mathcal{Q}_n A^*|_{\mathcal{X}_n \cap D(A^*)} : \mathcal{X}_n \cap D(A^*) \rightarrow \mathcal{X}_n^1$.

Remark 5.16. It can be proven that A_{12}^* from Definition 5.15 satisfies

$$(5.27) \quad \langle x, A_{12}y \rangle_{L^2} = \langle A_{12}^*x, y \rangle_{L^2} \quad \forall x \in \mathcal{X}_n \cap D(A^*), \forall y \in D(A),$$

explaining the notation. We do not need this fact for our argument and so omit the proof.

In the next lemma we collect several Lyapunov equations which the operators HH^* , A , B , C , A^* , and B^* and their truncations satisfy.

LEMMA 5.17. *Let H denote a Hankel operator satisfying (A) and let $L := HH^*$, which recall is a compact operator $L^1(\mathbb{R}^+; \mathcal{Y}) \rightarrow L^1(\mathbb{R}^+; \mathcal{Y})$, $L^2(\mathbb{R}^+; \mathcal{Y}) \rightarrow L^2(\mathbb{R}^+; \mathcal{Y})$, and $W^{1,1}(\mathbb{R}^+; \mathcal{Y}) \rightarrow W^{1,1}(\mathbb{R}^+; \mathcal{Y})$. Then L satisfies*

$$(5.28) \quad \langle x, Ly \rangle_{L^2} = \langle Lx, y \rangle_{L^2} \quad \forall x \in W^{1,1}(\mathbb{R}^+; \mathcal{Y}), \forall y \in L^1(\mathbb{R}^+; \mathcal{Y}),$$

and

$$(5.29) \quad \mathcal{Q}_n L|_{\mathcal{X}_n} = 0,$$

$$(5.30) \quad \mathcal{P}_n L|_{\mathcal{X}_n^1} = 0.$$

Define the decomposition

$$(5.31) \quad L = \begin{bmatrix} \mathcal{P}_n L|_{\mathcal{X}_n} & \mathcal{P}_n L|_{\mathcal{X}_n^1} \\ \mathcal{Q}_n L|_{\mathcal{X}_n} & \mathcal{Q}_n L|_{\mathcal{X}_n^1} \end{bmatrix} =: \begin{bmatrix} L_1 & 0 \\ 0 & L_2 \end{bmatrix}.$$

Let A, B, C, A^*, B^* denote the operators from Lemmas 5.1 and 5.3. Then the following equations hold on $D(A^*)$:

$$(5.32) \quad A^* + A = 0,$$

$$(5.33) \quad AL + LA^* + BB^* = 0,$$

and their related inner-product versions hold:

$$(5.34) \quad \langle Av, w \rangle_{L^2} + \langle v, Aw \rangle_{L^2} + \langle Cv, Cw \rangle_{\mathcal{Y}} = 0 \quad \forall v, w \in D(A),$$

$$(5.35) \quad \langle ALv, w \rangle_{L^2} + \langle v, ALw \rangle_{L^2} + \langle B^*v, B^*w \rangle_{\mathcal{Y}} = 0 \quad \forall v, w \in \mathcal{X}_n.$$

The L^2 inner products in the above two equations are understood as the duality pairing of L^1 and L^∞ (the latter containing $W^{1,1}$). The following truncated equations hold:

$$(5.36) \quad \langle A_{11}x, y \rangle_{L^2} + \langle x, A_{11}y \rangle_{L^2} + \langle C_1x, C_1y \rangle_{\mathcal{Y}} = 0 \quad \forall x, y \in \mathcal{X}_n,$$

$$(5.37) \quad \langle A_{11}L_1x, y \rangle_{L^2} + \langle x, A_{11}L_1y \rangle_{L^2} + \langle B_1^*x, B_1^*y \rangle_{\mathcal{Y}} = 0 \quad \forall x, y \in \mathcal{X}_n,$$

where A_{11}, B_1, B_1^*, C_1 are the operators from Remark 5.13 and Lemma 5.14. The following truncated operator equations hold on $\mathcal{X}_n \cap D(A^*)$:

$$(5.38) \quad A_{12}^* + A_{21} = 0,$$

$$(5.39) \quad A_{11}L_1 + L_1A_{11}^* + B_1B_1^* = 0,$$

$$(5.40) \quad A_{21}L_1 + L_2A_{12}^* + B_2B_1^* = 0.$$

Moreover, the following truncated operator equations hold on \mathcal{X}_n :

$$(5.41) \quad {}^2A_{11}^* + A_{11} + {}^2C_1^*C_1 = 0,$$

$$(5.42) \quad L_1 {}^2A_{11}^* + A_{11}L_1 + B_1B_1^* = 0.$$

The above operators are given by Remark 5.13, Lemma 5.14, Definition 5.15, and ${}^2C_1^* : \mathcal{Y} \rightarrow \mathcal{X}_n$ is the Hilbert space adjoint of C_1 .

Proof. The equalities are proven in order. Both sides of (5.28) make sense and are finite as

$$x \in W^{1,1} \Rightarrow Lx \in W^{1,1} \subseteq L^\infty \quad \text{and} \quad y \in L^1 \Rightarrow Ly \in L^1,$$

and so both sides of (5.28) are the pairing of an element of $W^{1,1}$ and an element of L^1 . To prove (5.28) let $x \in W^{1,1}$ and $y \in L^1$. Then as $W^{1,1}$ is dense in L^1 there exists a sequence $(y_m)_{m \in \mathbb{N}} \subseteq W^{1,1}$ such that

$$y_m \xrightarrow{L^1} y, \quad \text{as } m \rightarrow \infty.$$

By Remark 4.1 (ii), $W^{1,1} \subseteq L^2$ and so x and y_m are elements of L^2 . As $L : L^2 \rightarrow L^2$ is self-adjoint on L^2 ,

$$\langle x, Ly \rangle_{L^2} = \lim_{m \rightarrow \infty} \langle x, Ly_m \rangle_{L^2} = \lim_{m \rightarrow \infty} \langle Lx, y_m \rangle_{L^2} = \langle Lx, y \rangle_{L^2},$$

where we have used the continuity of L on L^1 and of the duality product.

We now prove (5.29) and (5.30). Observe that \mathcal{X}_n is the sum of the eigenspaces of L corresponding to the first n eigenvalues, so is L -invariant and thus (5.29) holds.

To prove (5.30) consider $x \in \mathcal{X}_n$ and $y \in \mathcal{X}_n^\perp$, so that $\mathcal{Q}_n y = y$ and thus

$$\begin{aligned} \langle Lx, y \rangle_{L^2} &= \langle L|_{\mathcal{X}_n} x, \mathcal{Q}_n y \rangle_{L^2} = \langle \mathcal{Q}_n L|_{\mathcal{X}_n} x, y \rangle_{L^2} \quad \text{by (5.10),} \\ &= 0 \quad \text{by (5.29).} \end{aligned}$$

Furthermore, by the self-adjointness of L in (5.28),

$$0 = \langle Lx, y \rangle_{L^2} = \langle x, Ly \rangle_{L^2} = \langle \mathcal{P}_n x, L|_{\mathcal{X}_n^\perp} y \rangle_{L^2} = \langle x, \mathcal{P}_n L|_{\mathcal{X}_n^\perp} y \rangle_{L^2} \quad \text{by (5.10).}$$

Therefore $\mathcal{P}_n L|_{\mathcal{X}_n^\perp} y \in \mathcal{X}_n$ and from the above is orthogonal to \mathcal{X}_n . We infer that

$$\mathcal{P}_n L|_{\mathcal{X}_n^\perp} y = 0 \quad \forall y \in \mathcal{X}_n^\perp,$$

and so (5.30) holds.

We now prove the Lyapunov equations. Equation (5.32) is established trivially given the definition of A^* in Lemma 5.3. For (5.33), let $x \in D(A^*) = W_0^{1,1}$ and $t \in \mathbb{R}^+$, so that from the derivative formula (4.19) for H^*x and Hx we compute

$$\begin{aligned} (AL + LA^*)x(t) &= \frac{d}{dt}(HH^*x)(t) - H(H^*\dot{x})(t) \\ &= -h(t)(H^*x)(0) - H\left(\underbrace{\frac{d}{dt}H^*x + H^*\dot{x}}_{=-h^*(\cdot)x(0)}\right)(t) \\ &= -h(t)(H^*x)(0), \quad \text{as } x(0) = 0, \\ &= -(BB^*x)(t). \end{aligned}$$

To establish (5.34) let $v, w \in D(A) = W^{1,1}$ and $t \geq 0$, so that

$$\langle \dot{v}(t), w(t) \rangle_{\mathcal{Y}} + \langle v(t), \dot{w}(t) \rangle_{\mathcal{Y}} = \frac{d}{dt} \langle v(t), w(t) \rangle_{\mathcal{Y}}.$$

Integrating both sides over \mathbb{R}^+ and using the fundamental theorem of calculus gives

$$\langle Av, w \rangle_{L^2} + \langle v, Aw \rangle_{L^2} = -\langle v(0), w(0) \rangle_{\mathcal{Y}} = -\langle Cv, Cw \rangle_{\mathcal{Y}},$$

which we can rearrange to give (5.34). To prove (5.35), note that $w_{i,k}, w_{j,r} \in \mathcal{X}_n$ are eigenvectors of L and thus

$$(5.43) \quad \langle ALw_{i,k}, w_{j,l} \rangle_{L^2} + \langle w_{i,k}, ALw_{j,l} \rangle_{L^2} = \langle \sigma_i^2 \dot{w}_{i,k}, w_{j,l} \rangle_{L^2} + \langle w_{i,k}, \sigma_j^2 \dot{w}_{j,l} \rangle_{L^2}.$$

A calculation now shows that

$$(5.44) \quad \begin{aligned} \langle w_{i,k}, \sigma_j \dot{w}_{j,l} \rangle_{L^2} &= \left\langle w_{i,k}, \frac{d}{dt} H v_{j,l} \right\rangle_{L^2} = [\langle w_{i,k}(t), H v_{j,l}(t) \rangle_{\mathcal{H}}]_0^\infty - \langle \dot{w}_{i,k}, H v_{j,l} \rangle_{L^2} \\ &= -\langle w_{i,k}(0), H v_{j,l}(0) \rangle_{\mathcal{H}} - \langle H^* \dot{w}_{i,k}, v_{j,l} \rangle_{L^2} \\ &= -\langle h^*(\cdot) w_{i,k}(0), v_{j,l} \rangle_{L^2} - \langle H^* \dot{w}_{i,k}, v_{j,l} \rangle_{L^2} \\ &= \langle -h^*(\cdot) w_{i,k}(0) - H^* \dot{w}_{i,k}, v_{j,l} \rangle_{L^2} \end{aligned}$$

$$(5.45) \quad = \left\langle \frac{d}{dt} H^* w_{i,k}, v_{j,l} \right\rangle_{L^2},$$

$$(5.45) \quad = \sigma_i \langle \dot{v}_{i,k}, w_{j,l} \rangle_{L^2},$$

where we have used the derivative formula for H^* in (5.44), which is given by (4.19) with h and H replaced by h^* and H^* , respectively. Multiplying both sides of (5.45) by σ_j implies that

$$(5.46) \quad \langle w_{i,k}, \sigma_j^2 \dot{w}_{j,l} \rangle_{L^2} = \sigma_i \sigma_j \langle \dot{v}_{i,k}, v_{j,l} \rangle_{L^2}.$$

Interchanging the indices i and j and k and l , the same calculation to that above gives

$$(5.47) \quad \langle \sigma_i^2 \dot{w}_{i,k}, w_{j,l} \rangle_{L^2} = \sigma_i \sigma_j \langle v_{i,k}, \dot{v}_{j,l} \rangle_{L^2}.$$

Substituting (5.46) and (5.47) into (5.43) yields that

$$(5.48) \quad \begin{aligned} \langle ALw_{i,k}, w_{j,l} \rangle_{L^2} + \langle w_{i,k}, ALw_{j,l} \rangle_{L^2} &= \sigma_i \sigma_j (\langle v_{i,k}, \dot{v}_{j,l} \rangle_{L^2} + \langle \dot{v}_{i,k}, v_{j,l} \rangle_{L^2}) \\ &= \int_{\mathbb{R}^+} \frac{d}{dt} \langle \sigma_i v_{i,k}(t), \sigma_j v_{j,l}(t) \rangle_{\mathcal{H}} dt \\ &= -\langle \sigma_i v_{i,k}(0), \sigma_j v_{j,l}(0) \rangle_{\mathcal{H}}, \\ &= -\langle (H^* w_{i,k})(0), (H^* w_{j,l})(0) \rangle_{\mathcal{H}} \\ &= -\langle B^* w_{i,k}, B^* w_{j,l} \rangle_{\mathcal{H}}. \end{aligned}$$

Equation (5.35) now follows by noting that any $x \in \mathcal{X}_n$ can be expressed as a linear combination of finitely many $w_{i,k}$, and that (5.48) is sesquilinear.

To prove (5.36) we start from (5.34), considered for $x, y \in \mathcal{X}_n \subseteq D(A)$,

$$\begin{aligned} 0 &= \langle Ax, y \rangle_{L^2} + \langle x, Ay \rangle_{L^2} + \langle Cx, Cy \rangle_{\mathcal{H}} \\ &= \langle A|_{\mathcal{X}_n} x, \mathcal{P}_n y \rangle_{L^2} + \langle \mathcal{P}_n x, A|_{\mathcal{X}_n} y \rangle_{L^2} + \langle C|_{\mathcal{X}_n} x, C|_{\mathcal{X}_n} y \rangle_{\mathcal{H}} \\ &= \langle \mathcal{P}_n A|_{\mathcal{X}_n} x, y \rangle_{L^2} + \langle x, \mathcal{P}_n A|_{\mathcal{X}_n} y \rangle_{L^2} + \langle C|_{\mathcal{X}_n} x, C|_{\mathcal{X}_n} y \rangle_{\mathcal{H}} \quad \text{by (5.10),} \end{aligned}$$

which is (5.36). The proof of (5.37) is similar to that above, starting from (5.35).

To prove (5.38) we apply \mathcal{Q}_n to (5.32), and consider for $x \in D := \mathcal{X}_n \cap D(A^*)$

$$0 = \mathcal{Q}_n(A^* + A)x = \mathcal{Q}_n A^*|_D x + \mathcal{Q}_n A|_{\mathcal{X}_n} x = A_{12}^* x + A_{21} x,$$

where we have used Definition 5.15 for A_{12}^* .

Next, for $x \in D$, applying \mathcal{P}_n to (5.33) gives

$$(5.49) \quad 0 = \mathcal{P}_n(AL + LA^* + BB^*)x = \mathcal{P}_nAL|_{\mathcal{X}_n}x + \mathcal{P}_nLA^*|_Dx + B_1B_1^*x.$$

We consider the first two terms on the right-hand side of (5.49) separately. First,

$$(5.50) \quad \begin{aligned} \mathcal{P}_nLA^*|_D &= \mathcal{P}_nL(\mathcal{P}_n + \mathcal{Q}_n)A^*|_D = \underbrace{\mathcal{P}_nL|_{\mathcal{X}_n}}_{=L_1} \mathcal{P}_nA^*|_D + \underbrace{\mathcal{P}_nL|_{\mathcal{X}_n^{1,1}}}_{=0 \text{ by (5.30)}} \mathcal{Q}_nA^*|_D \\ &= L_1^1A_{11}^* = L_1A_{11}^*. \end{aligned}$$

Second,

$$(5.51) \quad \mathcal{P}_nAL|_{\mathcal{X}_n} = \mathcal{P}_nA(\mathcal{P}_n + \mathcal{Q}_n)L|_{\mathcal{X}_n} = \mathcal{P}_nA|_{\mathcal{X}_n} \mathcal{P}_nL|_{\mathcal{X}_n} + \mathcal{P}_nA|_{\mathcal{X}_n^{1,1}} \mathcal{Q}_nL|_{\mathcal{X}_n},$$

where in (5.51) we have used that L maps $W^{1,1}$ into $W^{1,1}$ so that the compositions

$$\mathcal{P}_nL|_{\mathcal{X}_n} \quad \text{and} \quad \mathcal{Q}_nL|_{\mathcal{X}_n}$$

make sense. Now $\mathcal{P}_nL|_{\mathcal{X}_n}$ and L_1 are equal on \mathcal{X}_n , as \mathcal{P}_n and \mathcal{P}_n are equal on $\mathcal{X}_n = L\mathcal{X}_n$. Additionally, $\mathcal{Q}_nL|_{\mathcal{X}_n}$ is equal to $\mathcal{Q}_nL|_{\mathcal{X}_n}$ on \mathcal{X}_n , which is the zero map, and so $\mathcal{Q}_nL|_{\mathcal{X}_n}$ is also zero. Therefore, (5.51) becomes

$$(5.52) \quad \mathcal{P}_nAL|_{\mathcal{X}_n} = \mathcal{P}_nA|_{\mathcal{X}_n} \mathcal{P}_nL|_{\mathcal{X}_n} = A_{11}L_1.$$

Combining (5.49), (5.50), and (5.52) gives (5.39).

The proof of (5.40) is very similar to that of (5.39) only, instead, now we multiply (5.33) by \mathcal{Q}_n instead of \mathcal{P}_n . The Lyapunov equations (5.41)–(5.42) follow immediately from the inner-product versions (5.36) and (5.37), respectively, where in the second equation we have used the adjoint property of B_1^* in (5.23). \square

Proof of Proposition 5.12. We recap that we need to prove that the system $\begin{bmatrix} A_n & B_n \\ C_n & 0 \end{bmatrix}$ is stable, minimal, and output normal. These claims will follow in light of (5.41)–(5.42) (where $A_n = A_{11}$, $B_n = B_1$, $C_n = C_1$) once we establish that A_{11} is stable. In particular, if A_{11} is stable then from (5.41) it follows that the reduced order system is output normal. Output-normal realizations are trivially observable. Moreover, if A_{11} is stable then (5.42) implies that L_1 is the controllability Gramian of the reduced order system. The decomposition (5.31) demonstrates that with respect to the orthonormal basis $(w_{i,k})_{\substack{1 \leq k \leq p_i \\ 1 \leq i \leq n}}$ for \mathcal{X}_n , L_1 has matrix representation

$$\text{diag} \{ \sigma_1^2 I_{p_1}, \dots, \sigma_n^2 I_{p_n} \}, \quad I_p \text{ the identity matrix on } \mathbb{C}^p,$$

which is positive definite and thus the reduced order system is controllable. Additionally, the Hankel singular values of the reduced order system are the first n singular values of H , with the same multiplicities.

We therefore concentrate on proving the stability of A_{11} . As already mentioned, the argument that A_{11} is stable is based on the argument for finite-dimensional Lyapunov balanced truncation, but we need to be careful about which spaces the operators involved are defined on.

A short calculation using (5.36) demonstrates that every eigenvalue of A_{11} has a nonpositive real part. Seeking a contradiction, therefore, we assume that A_{11} has a purely imaginary eigenvalue λ . Let $Z \subseteq \mathcal{X}_n$ denote the eigenspace of A_{11} corresponding to λ . We observe immediately from (5.36) that for $x \in Z$

$$\langle C_1x, C_1x \rangle_{\mathcal{Y}} = -2\text{Re} \langle A_{11}x, x \rangle_{L^2} = -2\|x\|_2^2 (\text{Re } \lambda) = 0 \quad \Rightarrow \quad C_1x = 0$$

or, equivalently, C restricted to Z is zero. Since $Cz = z(0)$ we infer that

$$(5.53) \quad z \in Z \Rightarrow Cz = z(0) = 0 \Rightarrow Z \subseteq D(A^*) = W_0^{1,1}(\mathbb{R}^+; \mathcal{Y}),$$

in particular, $\{0\} \neq Z \subseteq \mathcal{X}_n \cap D(A^*)$. Considering (5.36) again for $x \in Z$ and $y \in \mathcal{X}_n$ and using (5.53) we observe that

$$0 = \langle A_{11}x, y \rangle_{L^2} + \langle x, A_{11}y \rangle_{L^2} = \langle \lambda x, y \rangle_{L^2} + \langle {}^2A_{11}^*x, y \rangle_{L^2} = \langle (\lambda I + {}^2A_{11}^*)x, y \rangle_{L^2},$$

and as $y \in \mathcal{X}_n$ was arbitrary,

$${}^2A_{11}^*x = -\lambda x = \bar{\lambda}x \quad \forall x \in Z.$$

Since $Z \subseteq \mathcal{X}_n \cap D(A^*)$, from Lemma 5.14 we see that ${}^1A_{11}^*$ and ${}^2A_{11}^*$ are equal on Z and so

$$(5.54) \quad A_{11}^*x = {}^1A_{11}^*x = {}^2A_{11}^*x = -\lambda x \quad \forall x \in Z.$$

For $x \in Z$, by using the adjoint A_{11}^* property in the Lyapunov equation (5.37) we obtain

$$\langle L_1x, A_{11}^*x \rangle_{L^2} + \langle A_{11}^*x, L_1x \rangle_{L^2} + \langle B_1^*x, B_1^*x \rangle_{\mathcal{U}} = 0,$$

which when we rearrange and use (5.54) yields

$$\begin{aligned} \langle B_1^*x, B_1^*x \rangle_{\mathcal{U}} &= -(\langle L_1x, A_{11}^*x \rangle_{L^2} + \langle A_{11}^*x, L_1x \rangle_{L^2}) = -2(\operatorname{Re} \lambda) \langle L_1x, x \rangle_{L^2} = 0, \\ \Rightarrow B_1^*x &= 0. \end{aligned}$$

We conclude that B^* restricted to Z is zero. Therefore from the truncated equation (5.39) we obtain for $x \in Z$

$$(5.55) \quad (A_{11}L_1 + L_1A_{11}^*)x = 0.$$

Inserting (5.54) into (5.55) gives $A_{11}(L_1x) = \lambda(L_1x)$ for any $x \in Z$, and so we infer that Z is L_1 -invariant. Now the truncated equation (5.38) yields for $x \in Z$

$$A_{12}^*x = -A_{21}x,$$

which when substituted into (5.40) gives

$$A_{21}L_1x + L_2A_{12}^*x + \underbrace{B_2B_1^*x}_{=0} = 0,$$

and so

$$(5.56) \quad A_{21}L_1 = L_2A_{21} \quad \text{on } Z.$$

Since Z is L_1 -invariant we can restrict L_1 to an operator $L_1^r : Z \rightarrow Z$, and remark that the spectrum of L_1^r is contained within the spectrum of L_1 . Let μ denote an eigenvalue of L_1^r , with corresponding eigenvector v . From (5.56) we note that

$$L_2(A_{21}v) = A_{21}L_1v = \mu(A_{21}v).$$

As L_1^r and L_2 have disjoint spectra, we conclude that $A_{21}v = 0$. Therefore, the operator A has eigenvector $v \in Z$, corresponding to the eigenvalue λ , as

$$Av = A|_{\mathcal{X}_n}v = \mathcal{P}_n A|_{\mathcal{X}_n}v + \mathcal{Q}_n A|_{\mathcal{X}_n}v = A_{11}v + A_{21}v = \lambda v.$$

As such the semigroup $\mathbb{S}(t)$ has eigenvector v with eigenvalue $e^{\lambda t}$. Recall that the realization (5.2) has as output map the identity, and so using (5.53) we obtain the contradiction

$$v(t) = (Iv)(t) = C\mathbb{S}(t)v = e^{\lambda t}Cv = 0.$$

We conclude that $A_n = A_{11}$ is stable and the proof is complete. \square

5.4. Relation to earlier work. We make comparisons between the results of section 5 and [13, sections 3 and 4]. Lemma 4.11 demonstrates that $h \in L^2$ occurs if and essentially only if the Schmidt pairs of H belong to $W^{1,2}$. The space $W^{1,2}$ is the domain of the main operator of the output-normal shift realization on L^2 ; the realization used in [13]. Without the assumption $h \in L^2$ we instead have to rely on Lemma 4.9 which states that $h \in L^1$ implies that the corresponding Schmidt pairs belong to $W^{1,1}$. As we seek realizations where we can naturally describe the truncations in terms of the Schmidt vectors we choose to use the exactly observable shift realization on L^1 . This is the L^1 equivalent of the output-normal realization (which is a Hilbert space concept).

Despite the different realization of the infinite-dimensional system used, our definition of reduced order system obtained by balanced truncation agrees with that of [13, section 4] in the sense that there the truncation is defined in terms of the matrices given by (5.17). Recall that these are (5.14) once adjusted for multiplicities of the singular values.

As Proposition 5.12 shows, the reduced order system we obtain is output normal and minimal. Moreover, as Theorem 2.3 demonstrates, using this truncation method we obtain the infinite-dimensional version of the Lyapunov balanced truncation error bound. Neither of these results require the extra assumptions of [13]. For instance, conclusion (2) of Proposition 5.10 is the same as [13, Lemma 4.5]. The key results there are [13, Lemma 4.3] and [13, Lemma 4.4]. The former establishes convergence of the Schmidt pairs in L^∞ and so also at zero and is proven by approximating $h \in L^2$ by h_m such that $h_m \rightarrow h$ in L^1 and L^2 as $m \rightarrow \infty$ (see Remark 4.14). Convergence of the Schmidt pairs at zero then gives componentwise convergence of the matrices \mathcal{B}_i^m and \mathcal{C}_i^m in (5.17) to \mathcal{B}_i and \mathcal{C}_i , respectively. The assumption that h is real or \dot{h} exists and is the kernel of a bounded Hankel operator is used in [13, Lemma 4.4] to prove componentwise convergence of the \mathcal{A}_{ij}^m . This assumption is unnecessary, and is avoided by using the $W^{1,1}$ convergence of Schmidt pairs from Theorem 4.4.

Throughout section 5 we did not need to assume that H is nuclear, only that assumption (A) holds.

6. Proof of Theorem 2.3. The proof is similar to that of [13, Theorem 5.1], only the technical results of [13] have been replaced with ours to accommodate our weaker assumptions. Specifically, [13, Lemma 4.4] has been replaced by Proposition 5.10. We also need to account for the multiplicities of the singular values, but note that the singular values are not repeated in the error bound according to multiplicity, in other words, the distinct singular values appear in the error bound (which is also the case for the finite-dimensional bound).

Proof of Theorem 2.3. We assume that G is irrational, which combined with the assumption that the Hankel operator H is nuclear, implies that the sequence of distinct singular values $(\sigma_j)_{j \in \mathbb{N}}$ has infinitely many nonzero (and so strictly positive) terms. There is no loss in generality in making such an assumption as the result is known when G is rational. We first prove that

$$(6.1) \quad \|G - G_n\|_{H^\infty} \rightarrow 0, \quad \text{as } n \rightarrow \infty,$$

by using partial sums of the Coifman and Rochberg decompositions of G from Proposition 3.4 as intermediary terms in (6.1) and applying Proposition 5.10. That is, for $m \in \mathbb{N}$ define

$$(6.2) \quad h_m(t) := \sum_{j=1}^m \lambda_j (\operatorname{Re} a_j) e^{a_j t}, \quad t > 0, \quad G^m(s) := \sum_{j=1}^m \lambda_j \frac{\operatorname{Re} a_j}{s - a_j}, \quad \operatorname{Re} s > 0,$$

so that the sequence $(H_m)_{m \in \mathbb{N}}$ given by (4.2) converges in nuclear norm to H . The following chain of inequalities holds:

$$(6.3) \quad \|G - G^m\|_{H^\infty} \leq \|h - h_m\|_1 \leq 2\|H - H_m\|_N,$$

where the second inequality is (3.4).

Note that as G^m is rational for each $m \in \mathbb{N}$, its Hankel operator H_m is finite rank. Therefore, H_m has only finitely many distinct singular values, the number of which we denote by $\mathcal{N}(m)$. As usual we denote by $(\sigma_j^{(m)})_{j=1}^{\mathcal{N}(m)}$ the distinct singular values of H_m . From our assumption that H has infinitely many nonzero, distinct singular values, the convergence (4.3) in Theorem 4.4 implies that $(\mathcal{N}(m))_{m \in \mathbb{N}}$ is unbounded from above. Thus for fixed $n \in \mathbb{N}$, there exists M_1 (which depends on n) such that

$$(6.4) \quad m \in \mathbb{N} \quad \text{and} \quad m \geq M_1 \quad \Rightarrow \quad \mathcal{N}(m) \geq l_n + 1 \geq n + 1,$$

where $(l_n)_{n \in \mathbb{N}}$ is the increasing sequence of positive integers from Theorem 4.4 which, recall, satisfy $l_n \geq n$ for each $n \in \mathbb{N}$. For $m, n \in \mathbb{N}$ we let $G_{l_n}^m$ denote the balanced truncation of G^m which, as described in Remark 5.6(ii), is well-defined whenever $\mathcal{N}(m) \geq l_n$. By (6.4) we have that $G_{l_n}^m$ is well-defined whenever $m \geq M_1$.

Let $\varepsilon > 0$ be given. Since H is nuclear, there exists $N \in \mathbb{N}$ such that

$$(6.5) \quad n \in \mathbb{N} \quad \text{and} \quad n \geq N \quad \Rightarrow \quad \sum_{k=n+1}^{\infty} p_k \sigma_k < \frac{\varepsilon}{16}.$$

Now fix $n \geq N$ and assume that $l_n \geq n + 1$. The alternative case where $l_n = n$ will be addressed later. The Lyapunov balanced truncation error bound for rational transfer functions applies for estimating the difference $G_{l_n} - G_n$, namely,

$$(6.6) \quad \|G_{l_n} - G_n\|_{H^\infty} \leq 2 \sum_{k=n+1}^{l_n} \sigma_k \leq 2 \sum_{k=n+1}^{\infty} p_k \sigma_k < \frac{\varepsilon}{8},$$

where we have used the bound (6.5). For the above we have used that the output-normal realization $\begin{bmatrix} A_n & B_n \\ C_n & 0 \end{bmatrix}$ of G_n is the balanced truncation of the output-normal realization $\begin{bmatrix} A_{l_n} & B_{l_n} \\ C_{l_n} & 0 \end{bmatrix}$ of G_{l_n} , which follows from Proposition 5.12.

By choice of the H_m , we can choose $M_2 \in \mathbb{N}$, independently of n , such that

$$(6.7) \quad m \in \mathbb{N} \quad \text{and} \quad m \geq M_2 \quad \Rightarrow \quad \|H - H_m\|_N < \frac{\varepsilon}{16},$$

so that by (6.3) for $m \geq M_2$

$$(6.8) \quad \|G - G^m\|_{H^\infty} < \frac{\varepsilon}{8}.$$

Now choose $M_3 \in \mathbb{N}$ (which depends on n) with $M_3 \geq M_1$ such that

$$(6.9) \quad m \in \mathbb{N} \quad \text{and} \quad m \geq M_3 \quad \Rightarrow \quad \left| \sum_{j=1}^{l_n} p_j^{(m)} \sigma_j^{(m)} - \sum_{j=1}^n p_j \sigma_j \right| < \frac{\varepsilon}{16},$$

which is possible by the convergence in (4.3). For each $m \in \mathbb{N}$ with $m \geq M_3$ we now invoke the Lyapunov balanced truncation error bound for rational transfer functions again to estimate the difference $G^m - G_{l_n}^m$, namely,

$$(6.10) \quad \|G^m - G_{l_n}^m\|_{H^\infty} \leq 2 \sum_{k=l_n+1}^{\mathcal{N}(m)} \sigma_k^{(m)}.$$

We use the bound (6.9) to show that for $m \geq M_3$ the right-hand side of (6.10) can be bounded by an arbitrarily small term. Specifically,

$$\begin{aligned} \sum_{k=l_n+1}^{\mathcal{N}(m)} \sigma_k^{(m)} &\leq \sum_{k=l_n+1}^{\mathcal{N}(m)} p_k^{(m)} \sigma_k^{(m)} = \left(\sum_{k=l_n+1}^{\mathcal{N}(m)} p_k^{(m)} \sigma_k^{(m)} - \sum_{k=1}^{\infty} p_k \sigma_k \right) + \sum_{k=1}^{\infty} p_k \sigma_k \\ &= \left(\sum_{k=1}^{\mathcal{N}(m)} p_k^{(m)} \sigma_k^{(m)} - \sum_{k=1}^{\infty} p_k \sigma_k \right) + \sum_{k=n+1}^{\infty} p_k \sigma_k \\ &\quad - \left(\sum_{k=1}^{l_n} p_k^{(m)} \sigma_k^{(m)} - \sum_{k=1}^n p_k \sigma_k \right). \end{aligned}$$

Therefore, by the triangle inequality

$$(6.11) \quad \begin{aligned} 2 \left| \sum_{k=l_n+1}^{\mathcal{N}(m)} \sigma_k^{(m)} \right| &\leq 2 \left(\|H - H_m\|_N + \sum_{k=n+1}^{\infty} p_k \sigma_k + \left| \sum_{k=1}^{l_n} p_k^{(m)} \sigma_k^{(m)} - \sum_{k=1}^n p_k \sigma_k \right| \right) \\ &< 2 \left(\frac{\varepsilon}{16} + \frac{\varepsilon}{16} + \frac{\varepsilon}{16} \right) = \frac{3\varepsilon}{8}, \end{aligned}$$

where we have obtained (6.11) by appealing to the bounds (6.7), (6.5), and (6.9), respectively. Combining (6.10) and (6.11) we obtain for $m \geq M_3$

$$(6.12) \quad \|G^m - G_{l_n}^m\|_{H^\infty} < \frac{3\varepsilon}{8}.$$

The inequalities (6.3) imply that the impulse responses h_m converge to h in L^1 as $m \rightarrow \infty$ and so the conditions of Proposition 5.10 are satisfied. By this result we can

choose $M_4 \in \mathbb{N}$ (which depends on n) with $M_4 \geq M_1$ and a subsequence $(\tau(m))_{m \in \mathbb{N}}$ such that

$$(6.13) \quad m \in \mathbb{N} \quad \text{and} \quad \tau(m) \geq m \geq M_4 \quad \Rightarrow \quad \|G_{l_n}^{\tau(m)} - G_{l_n}\|_{H^\infty} < \frac{3\varepsilon}{8}.$$

It remains to combine the bounds (6.6), (6.8), (6.12), and (6.13). Choose $m \in \mathbb{N}$ such that $\tau(m) \geq m \geq \max\{M_2, M_3, M_4\} \geq M_1$ from which we estimate

$$(6.14) \quad \begin{aligned} \|G - G_n\|_{H^\infty} &\leq \|G - G^{\tau(m)}\|_{H^\infty} + \|G^{\tau(m)} - G_{l_n}^{\tau(m)}\|_{H^\infty} \\ &\quad + \|G_{l_n}^{\tau(m)} - G_{l_n}\|_{H^\infty} + \|G_{l_n} - G_n\|_{H^\infty} \end{aligned}$$

$$(6.15) \quad < \frac{\varepsilon}{8} + \frac{3\varepsilon}{8} + \frac{3\varepsilon}{8} + \frac{\varepsilon}{8} = \varepsilon.$$

To summarize, for every $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that $n \in \mathbb{N}$ and $n \geq N$ implies that (6.15) holds; equivalently, we have proven (6.1). If $l_n = n$ then the proof is the same as above, only noting that here $G_{l_n} = G_n$, so although the bound (6.6) no longer makes sense, the final term on the right-hand side of (6.14) is zero and the proof of (6.12) is as before.

To prove the error bound (2.4) for $n \in \mathbb{N}$ we use the finite-dimensional Lyapunov balanced truncation error bound

$$(6.16) \quad \|G_j - G_n\|_{H^\infty} \leq 2 \sum_{k=n+1}^j \sigma_k \leq 2 \sum_{k=n+1}^\infty \sigma_k, \quad j > n.$$

Let $\varepsilon > 0$ be given so that by (6.1) we can choose $j \in \mathbb{N}$, $j > n$ such that

$$\|G - G_n\|_{H^\infty} \leq \|G - G_j\|_{H^\infty} + \|G_j - G_n\|_{H^\infty} \leq \varepsilon + 2 \sum_{k=n+1}^\infty \sigma_k,$$

where we have used (6.16) to bound the second term above. Since $\varepsilon > 0$ was arbitrary, we conclude that (2.4) holds. \square

REFERENCES

- [1] V. M. ADAMJAN, D. Z. AROV, AND M. G. KREĬN, *Analytic properties of the Schmidt pairs of a Hankel operator and the generalized Schur-Takagi problem*, Mat. Sb., 86 (1971), pp. 34–75.
- [2] C. BONNET, *Approximation of infinite-dimensional discrete time linear systems via balanced realizations and an application to fractional filters*, in Analysis and Optimization of Systems: State and Frequency Domain Approaches for Infinite-Dimensional Systems, Lecture Notes in Control and Inform. Sci. 185, Springer, Berlin, 1993, pp. 574–584.
- [3] C. BONNET, *Convergence and convergence rate of the balanced realization truncations for infinite-dimensional discrete-time systems*, Systems Control Lett., 20 (1993), pp. 353–359.
- [4] F. CHATELIN, *Spectral Approximation of Linear Operators*, Comput. Sci. Appl. Math., Academic, New York, 1983.
- [5] R. R. COIFMAN AND R. ROCHBERG, *Representation theorems for holomorphic and harmonic functions in L^p* , in Representation Theorems for Hardy Spaces, Astérisque 77, Société Mathématique de France, Paris, 1980, pp. 11–66.
- [6] R. F. CURTAIN AND K. GLOVER, *Balanced realisations for infinite-dimensional systems*, in Operator Theory and Systems, Oper. Theory Adv. Appl. 19, Birkhäuser, Basel, 1986, pp. 87–104.
- [7] R. F. CURTAIN AND A. J. SASANE, *Compactness and nuclearity of the Hankel operator and internal stability of infinite-dimensional state linear systems*, Internat. J. Control, 74 (2001), pp. 1260–1270.

- [8] R. F. CURTAIN AND H. ZWART, *An Introduction to Infinite-Dimensional Linear Systems Theory*, Texts Appl. Math. 21, Springer-Verlag, New York, 1995.
- [9] U. B. DESAI AND D. PAL, *A transformation approach to stochastic model reduction*, IEEE Trans. Automat. Control, 29 (1984), pp. 1097–1100.
- [10] D. F. ENNS, *Model reduction with balanced realizations: An error bound and a frequency weighted generalization*, in Proceedings of the 23rd IEEE Conference on Decision and Control, Las Vegas, 1984, IEEE, New York, 1984, pp. 127–132.
- [11] K. GLOVER, *All optimal Hankel-norm approximations of linear multivariable systems and their L^∞ -error bounds*, Internat. J. Control, 39 (1984), pp. 1115–1193.
- [12] K. GLOVER, *A tutorial on Hankel-norm approximation*, in From Data to Model, Springer, Berlin, 1989, pp. 26–48.
- [13] K. GLOVER, R. F. CURTAIN, AND J. R. PARTINGTON, *Realisation and approximation of linear infinite-dimensional systems with error bounds*, SIAM J. Control Optim., 26 (1988), pp. 863–898.
- [14] K. GLOVER, J. LAM, AND J. R. PARTINGTON, *Rational approximation of a class of infinite-dimensional systems. I. Singular values of Hankel operators*, Math. Control Signals Systems, 3 (1990), pp. 325–344.
- [15] I. GOHBERG, S. GOLDBERG, AND M. A. KAASHOEK, *Classes of Linear Operators*, Vol. I, Oper. Theory Adv. Appl. 49, Birkhäuser Verlag, Basel, 1990.
- [16] M. GREEN AND D. J. N. LIMEBEER, *Linear Robust Control*, Prentice-Hall, Upper Saddle River, NJ, 1995.
- [17] C. GUIVER, *Model Reduction by Balanced Truncation*, Ph.D. thesis, University of Bath, Bath, UK, 2012.
- [18] M. B. HA, *Model Reduction in a Behavioral Framework*, Ph.D. thesis, University of Groningen, Groningen, The Netherlands, 2009.
- [19] P. D. LAX, *Symmetrizable linear transformations*, Comm. Pure Appl. Math., 7 (1954), pp. 633–647.
- [20] P. D. LAX, *Functional Analysis*, Pure Appl. Math., John Wiley, New York, 2002.
- [21] B. C. MOORE, *Principal component analysis in linear systems: Controllability, observability, and model reduction*, IEEE Trans. Automat. Control, 26 (1981), pp. 17–32.
- [22] K. A. MORRIS, *Design of finite-dimensional controllers for infinite-dimensional systems by approximation*, J. Math. Systems Estim. Control, 4 (1994).
- [23] N. K. NIKOLSKI, *Operators, Functions, and Systems: An Easy Reading*, Vol. 1, Math. Surveys Monogr. 92, AMS, Providence, RI, 2002.
- [24] R. OBER AND S. MONTGOMERY-SMITH, *Bilinear transformation of infinite-dimensional state-space systems and balanced realizations of nonrational transfer functions*, SIAM J. Control Optim., 28 (1990), pp. 438–465.
- [25] P. C. OPDENACKER AND E. A. JONCKHEERE, *A contraction mapping preserving balanced reduction scheme and its infinity norm error bounds*, IEEE Trans. Circuits Systems, 35 (1988), pp. 184–189.
- [26] M. R. OPMEER, *Model Reduction for Controller Design for Infinite-Dimensional Systems*, Ph.D. thesis, University of Groningen, Groningen, The Netherlands, 2006.
- [27] M. R. OPMEER, *Decay of Hankel singular values of analytic control systems*, Systems Control Lett., 59 (2010), pp. 635–638.
- [28] L. B. PAGE, *Bounded and compact vectorial Hankel operators*, Trans. Amer. Math. Soc., 150 (1970), pp. 529–539.
- [29] J. R. PARTINGTON, *An Introduction to Hankel Operators*, London Math. Soc. Stud. Texts 13, Cambridge University Press, Cambridge, 1988.
- [30] V. V. PELLER, *Hankel Operators and Their Applications*, Springer Monogr. Math., Springer-Verlag, New York, 2003.
- [31] L. PERNEBO AND L. M. SILVERMAN, *Model reduction via balanced state space representations*, IEEE Trans. Automat. Control, 27 (1982), pp. 382–387.
- [32] T. REIS AND T. STYKEL, *Positive real and bounded real balancing for model reduction of descriptor systems*, Internat. J. Control, 83 (2010), pp. 74–88.
- [33] M. ROSENBLUM AND J. ROVNYAK, *Hardy classes and operator theory*, Dover, Mineola, NY, 1997.
- [34] D. SALAMON, *Infinite-dimensional linear systems with unbounded control and observation: A functional analytic approach*, Trans. Amer. Math. Soc., 300 (1987), pp. 383–431.
- [35] D. SALAMON, *Realization theory in Hilbert space*, Math. Systems Theory, 21 (1989), pp. 147–164.
- [36] J. R. SINGLER, *Balanced POD for model reduction of linear PDE systems: Convergence theory*, Numer. Math., 121 (2012), pp. 127–164.

- [37] O. J. STAFFANS, *Well-Posed Linear Systems*, Encyclopedia Math. Appl. 103, Cambridge University Press, Cambridge, 2005.
- [38] G. WEISS, *Admissibility of unbounded control operators*, SIAM J. Control Optim., 27 (1989), pp. 527–545.
- [39] G. WEISS, *Admissible observation operators for linear semigroups*, Israel J. Math., 65 (1989), pp. 17–43.
- [40] G. WEISS, *Transfer functions of regular linear systems. I. Characterizations of regularity*, Trans. Amer. Math. Soc., 342 (1994), pp. 827–854.
- [41] N. J. YOUNG, *Balanced realizations in infinite dimensions*, in Operator Theory and Systems, Oper. Theory Adv. Appl. 19, Birkhäuser, Basel, 1986, pp. 449–471.
- [42] K. ZHOU, J. C. DOYLE, AND K. GLOVER, *Robust and Optimal Control*, Prentice Hall, Englewood Cliffs, NJ, 1996.