SIAM J. CONTROL OPTIM.
Vol. 54, No. 6, pp. 2992–3015

STRUCTURE-PRESERVING MODEL REDUCTION FOR INTEGRO-DIFFERENTIAL EQUATIONS*

TOBIAS BREITEN†

Abstract. Model reduction for linear Volterra integro-diffferential equations is studied. Generalized system Gramians are introduced and characterized as solutions to delay Lyapunov equations similarly arising for finite delay systems. The usual energy interpretation of the Gramians is provided and a reduced-order model of Volterra integro-differential type is obtained by truncation of a balanced system. An error bound for the $\mathcal{H}_2$-norm is derived. It is further shown that particular choices for the Volterra kernel automatically yield approaches that have been studied in the literature. Additionally, the new approach allows us to also reduce time fractional systems. The method is numerically investigated by means of two spatially discretized partial differential equations.

Key words. integro-differential equations, model reduction, balanced truncation

AMS subject classifications. 93A15, 93A30, 93B40, 93C20

DOI. 10.1137/15M1032296

1. Introduction.

Consider a system of linear Volterra integro-differential equations of convolution type

$$\begin{aligned} \dot{x}(t) &= \int_0^t \mu(\mathrm{d}s)x(t-s) + \mathbf{B}u(t), \\ y(t) &= \mathbf{C}x(t), \quad x(0) = x_0, \end{aligned} \tag{1.1}$$

where $\mathbf{B} \in \mathbb{R}^{n\times m}, \mathbf{C} \in \mathbb{R}^{p\times n}$, and $\mu \in \mathcal{M}_{\mathrm{loc}}(0,\infty;\mathbb{R}^{n\times n})$ is a locally finite matrix valued Borel measure. In particular, we focus on the setup used in [19, Chapter 3] and assume that μ denotes a matrix in $\mathbb{R}^{n\times n}$ consisting of scalar measures. For fixed time t, we call $x(t) \in \mathbb{R}^n, u(t) \in \mathbb{R}^m, y(t) \in \mathbb{R}^p$ the *state*, *input*, and *output* of the system at time t, respectively.

Note that instead of (1.1), one often has to consider an additional mass matrix $\mathbf{E} \in \mathbb{R}^{n\times n}$ leading to systems of the form

$$\begin{aligned} \mathbf{E}\dot{x}(t) &= \int_0^t \mu(\mathrm{d}s)x(t-s) + \mathbf{B}u(t), \\ y(t) &= \mathbf{C}x(t), \quad x(0) = x_0. \end{aligned}$$

As long as the dynamics contain no (implicit) algebraic constraints such that $\mathbf{E}$ is singular, all of the following concepts can be appropriately extended. Indeed, for the theoretical results we might simply replace the previous system by the following one,

$$\begin{aligned} \dot{x}(t) &= \int_0^t \tilde{\mu}(\mathrm{d}s)x(t-s) + \tilde{\mathbf{B}}u(t), \\ y(t) &= \mathbf{C}x(t), \quad x(0) = x_0, \end{aligned}$$

*Received by the editors July 23, 2015; accepted for publication (in revised form) August 23, 2016; published electronically November 8, 2016.
http://www.siam.org/journals/sicon/54-6/M103229.html
†Institute for Mathematics and Scientic Computing, Karl-Franzens-Universität, Heinrichstr. 36, 8010 Graz, Austria (tobias.breiten@uni-graz.at).

where $\tilde{\mu} = \mathbf{E}^{-1}\mu$ and $\tilde{\mathbf{B}} = \mathbf{E}^{-1}\mathbf{B}$. On the other hand, in case of $\mathbf{E}$ being singular, we obtain an integro-differential algebraic system which requires many additional concepts. We thus restrict ourselves to the case of (1.1).

The study of equations of the form (1.1) has a long history going back to Volterra [45]. For a detailed introduction into the theory of integro-differential equations, we also refer to, e.g., [9, 19]. While several works in the literature [9, 20, 31, 34, 43] are concerned with the case $\mu(\mathrm{d}s) = k(s)\mathrm{d}s$, where $k(\cdot) \in L^1(0,\infty;\mathbb{R}^{n\times n})$, here we focus on the more general formulation (1.1). Note in particular that by setting $\mu(\mathrm{d}s) = \mathbf{A}\delta_0(\mathrm{d}s)$, where δ_0 is a scalar unit point mass at 0, we obtain a linear time invariant (LTI) control system of the form

$$\begin{aligned} \dot{x}(t) &= \mathbf{A}x(t) + \mathbf{B}u(t), \\ y(t) &= \mathbf{C}x(t), \quad x(0) = x_0. \end{aligned} \tag{1.2}$$

In what follows, we subsequently comment on that as well as other relations which allow for a comparison with existing results from the literature.

Often the need for an accurate model of the underlying physical process competes with the desire for an efficient numerical treatment of the mathematical system. As a remedy for this problem, the topic of *model order reduction* has gained much attention in recent years; see, e.g., [1, 5]. Given a system of the form (1.1), structure-preserving model reduction typically consists in constructing a *reduced-order model* of the same form, i.e.,

$$\begin{aligned} \dot{x}_r(t) &= \int_0^t \mu_r(\mathrm{d}s)x_r(t-s) + \mathbf{B}_r u(t), \\ y_r(t) &= \mathbf{C}_r x_r(t), \quad x_r(0) = x_{0,r}, \end{aligned} \tag{1.3}$$

where $\mathbf{B}_r \in \mathbb{R}^{r\times m}, \mathbf{C}_r \in \mathbb{R}^{p\times r}$, and $\mu_r \in \mathcal{M}_{\mathrm{loc}}(0,\infty;\mathbb{R}^{r\times r})$ such that $r \ll n$. Note that the term *structure-preserving refers to the fact that the reduced-order model is also of Volterra integro-differential type. We emphasize that there are alternative ways of defining what constitutes a reduced-order model for* (1.1). *From a system theoretic point of view, Volterra integro-differential systems are infinite dimensional; see subsection* 2.3. *Hence, instead of* (1.3) *one might also think of a model reduction problem where* (1.1) *is replaced by a standard LTI system of the form* (1.2).

In order to ensure now that the reduced model reflects the original dynamics, one also requires that $y_r(t) \approx y(t)$ for a given set of admissible input signals $u(t)$. Here, we assume that $u(\cdot) \in L^2(0,\infty;\mathbb{R}^m)$. As a measure for the deviation between original and reduced output, one may consider the $\mathcal{H}_\infty$-norm or the $\mathcal{H}_2$-norm of the systems; cf. [15, Appendix A.6]. For systems of the form (1.2), many different methodologies have been proposed and theoretically analyzed; among them the most popular ones can be sorted into balancing based methods [1, 5, 21, 32, 33], Krylov subspace or rational interpolation based methods [2, 17, 18, 22, 44], and methods based on proper orthogonal decomposition (POD) [12, 24, 29, 40, 41].

In this paper, a generalization of the method of balanced truncation to integro-differential equations of the form (1.1) is considered. The main results can also be seen as an appropriate extension of the method of *position balancing* for time delay systems [25] and *position-velocity balancing* for second order systems [11, 39]. In particular, both techniques can be obtained as special cases from the general framework presented below.

The precise structure of the paper is as follows. In section 2 we provide the necessary background for integro-differential equations of convolution type. Stability

criteria and the corresponding fundamental solution of the system are introduced. A generalized transfer function relating input and output signals in the frequency domain is studied. In section 3, we define two system Gramians that, in the case of (1.2), coincide with the solutions of the controllability and observability Lyapunov equations of the system. Moreover, they are shown to provide the usual energy interpretation of the states of (1.1) and can be characterized as the solution to a *delay Lyapunov equation*. The reduced model is obtained after a contragredient transformation and a truncation of the resulting balanced model. For the reduced-order model, an error bound for the $\mathcal{H}_2$-norm is derived. Section 4 contains a discussion of some special measures μ. Besides the already mentioned second order and time delay systems, we explain how the approach can be used to reduce time fractional as well as coupled systems. The numerical realization of the method is studied in section 5. The required Gramians are computed by means of a low-rank version of the Gauss–Kronrod quadrature formula. Reduced models are obtained for the linear heat equation with memory and a time fractional wave equation. A conclusion is given in section 6.

In the remainder of the paper, we use the following notation. All matrices are denoted in bold letters and $\mathbf{I}$ always denotes the identity matrix whose dimension should be clear from the context. Given a matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$, we denote with $\mathbf{A}^\dagger$ its Moore–Penrose pseudoinverse. For a matrix $\mathbf{U} \in \mathbb{R}^{n\times m}$, by $\mathbf{U}_{(:,1:r)} \in \mathbb{R}^{n\times r}$ we denote the submatrix consisting of the first r columns of the matrix $\mathbf{U}$. Similarly, we use $\mathbf{U}_{(1:s,1:r)} \in \mathbb{R}^{s\times r}$ for the submatrix consisting of the first s rows and the first r columns of the matrix $\mathbf{U}$. For $p \geq 1, k \in \mathbb{N}_0$, by $L^p(I)$ and $W^{k,p}(I)$ we denote the usual Lebesgue and Sobolev spaces on the interval $I \subset \mathbb{R}$. Given a Hilbert space X, we denote by $L^p(0,\infty;X)$ (Bochner) p-integrable functions on $(0,\infty)$ with values in X. The open left complex half-plane is denoted by $\mathbb{C}_- := \{s \in \mathbb{C} | \mathrm{Re}(s) < 0\}$. Similarly, we use $\overline{\mathbb{C}}_-, \mathbb{C}_+$, and $\overline{\mathbb{C}}_+$.

2. Integro-differential equations.

2.1. Preliminaries. For the results in this section, we closely follow the presentation in [19, Chapter 3]. Let us emphasize that this particularly applies to the solution concept of equations of the form (1.1). As pointed out in [19], even for continuous input signals $u(t)$, we cannot expect $x(t)$ to satisfy (1.1) everywhere. This is due to the fact that μ may consist of point masses and $\int_0^t \mu(\mathrm{d}s)x(t-s)$ will have jump discontinuities for which $\dot{x}(t)$ is not defined. Consequently, solutions are assumed to be locally absolutely continuous functions $x(t)$ with $x(0) = x_0$ such that (1.1) holds for almost all t. An important tool then is the so-called *differential resolvent* $\Phi(t)$.

THEOREM 2.1 (see [19, Theorem 3.3.1]). *Let $\mu \in \mathcal{M}_{\mathrm{loc}}(0,\infty;\mathbb{R}^{n\times n})$. Then there is a unique locally absolutely continuous function $\Phi(\cdot)$ on $[0,\infty)$ with $\Phi(0) = \mathbf{I}$ that satisfies*

$$\dot{\Phi}(t) = \int_0^t \mu(\mathrm{d}s)\Phi(t-s) = \int_0^t \Phi(t-s)\mu(\mathrm{d}s) \tag{2.1}$$

for almost all $t \in [0,\infty)$. Furthermore, the derivative $\dot{\Phi}$ is equal almost everywhere to a function which is locally of bounded variation.

For later purposes, we assume that Φ is also defined on $(-\infty, 0)$ with $\Phi(t) = 0$ for $t < 0$. The solution $x(t)$ of (1.1) is given by the *variation of constants formula* as

$$x(t) = \Phi(t)x_0 + \int_0^t \Phi(t-s)\mathbf{B}u(s)\,\mathrm{d}s. \tag{2.2}$$

Note that in the case of (1.2), the differential resolvent is the usual matrix exponential, i.e., $\Phi(t) = e^{\mathbf{A}t}$, which is also called a fundamental solution of the system.

For studying (1.1) in the frequency domain, we recapitulate the Laplace transform of measures.

DEFINITION 2.2 (see [19, Definition 3.2.2]). *The Laplace transform $\hat{\mu}(z)$ of a locally finite measure μ on $\mathbb{R}^+$ is the function*

$$\hat{\mu}(z) = \int_0^\infty e^{-zt}\mu(\mathrm{d}t),$$

defined for those $z \in \mathbb{C}$ for which this integral converges absolutely.

Since the Laplace transform of measures shares most of the properties of the Laplace transform of L^1-functions, we obtain a simple algebraic relation between inputs and outputs. Assuming that $x(0) = 0$ and applying the Laplace transform to (1.1) it holds that

$$\hat{y}(z) = \mathbf{C}(z\mathbf{I} - \hat{\mu}(z))^{-1}\mathbf{B}\hat{u}(z). \tag{2.3}$$

As a consequence, we will call

$$\mathbf{G}(z) := \mathbf{C}(z\mathbf{I} - \hat{\mu}(z))^{-1}\mathbf{B}$$

the *transfer function* of the system. Note that in contrast to systems of the form (1.2), $\mathbf{G}(z)$ is not necessarily a rational function in z.

In case of an initial condition $x(0) \neq 0$, we rather have to consider

$$\begin{aligned}\hat{y}(z) &= \mathbf{C}(z\mathbf{I} - \hat{\mu}(z))^{-1}\mathbf{B}\hat{u}(z) + \mathbf{C}(z\mathbf{I} - \hat{\mu}(z))^{-1}x(0)\\ &= \mathbf{C}(z\mathbf{I} - \hat{\mu}(z))^{-1}\underbrace{\begin{bmatrix}\mathbf{B} & x(0)\end{bmatrix}}_{\tilde{\mathbf{B}}}\underbrace{\begin{bmatrix}\hat{u}(z)\\ 1\end{bmatrix}}_{\tilde{u}(z)}.\end{aligned}$$

In other words, we can interpret the transformed transfer function as the transfer function of a system with zero initial condition and constant input term. Since this general case leads to additional difficulties [23], here we only consider the case $x(0) = 0$.

At this point, let us emphasize that instead of the balancing based model reduction approach we consider here, a very general interpolatory projection framework has been studied in [4]. In particular, it is shown how to construct a reduced-order model of the form (1.3) whose transfer function $\mathbf{G}_r$ interpolates the original transfer function $\mathbf{G}$ at a set of prescribed interpolation points.

2.2. Stability of Volterra equations and the $\mathcal{H}_2$-norm. For what follows, we first need an appropriate notion of stability for (1.1). Among many works in the literature [9, 20, 31, 34, 43], we give the following stability results from [19].

THEOREM 2.3 (see [19, Theorem 3.3.5]). *Let $\mu \in \mathcal{M}(0,\infty;\mathbb{R}^{n\times n})$. Then*

$$\Phi(\cdot) \in L^1(0,\infty;\mathbb{R}^{n\times n}) \Leftrightarrow \det(z\mathbf{I} - \hat{\mu}(z)) \neq 0,\ \mathrm{Re}(z) \geq 0.$$

Moreover, if $\Phi(\cdot) \in L^1(0,\infty;\mathbb{R}^{n\times n})$, then $\dot{\Phi} \in L^1(0,\infty;\mathbb{R}^{n\times n})$ and $\dot{\Phi}(\cdot)$ is equal almost everywhere to a function of bounded variation.

In particular, according to [19, Theorem 3.3.9], the previous result also implies that $\Phi(\cdot) \in L^p(0,\infty;\mathbb{R}^{n\times n})$ for all $p \in [1,\infty]$. For unbounded measures, one still has the following result.

THEOREM 2.4 (see [19, Theorem 3.3.13]). *Let $\mu \in \mathcal{M}_{\mathrm{loc}}(0,\infty;\mathbb{R}^{n\times n})$ satisfy $\int_{-\infty}^{\infty} e^{-\sigma t}|\mu|(\mathrm{d}t) < \infty$ for all $\sigma > 0$. Then the following are equivalent:*

(1) $\Phi(\cdot) \in L^2(0,\infty;\mathbb{R}^{n\times n})$,

(2) $\det(z\mathbf{I} - \hat{\mu}(z)) \neq 0,\ \mathrm{Re}(z) > 0,\ \sup_{\sigma>0} \int_{-\infty}^{\infty} \left\| ((\sigma + \imath\omega)\mathbf{I} - \hat{\mu}(\sigma+\imath\omega))^{-1} \right\|^2 \mathrm{d}\omega < \infty.$

With this in mind, throughout the rest of the paper, we assume that system (1.1) is stable in the sense that it at least satisfies the asymptotic properties stated in Theorem 2.4. In case of (1.2), it is well known [1, 15] that the second characterization in Theorem 2.4 is closely related to the *Hardy space* $\mathcal{H}_2^{p\times m}$ defined as follows,

$$\mathcal{H}_2^{p\times m} := \left\{ f\colon \mathbb{C}_0^+ \to \mathbb{C}^{p\times m} | f \text{ is holomorphic, and } \sup_{\sigma>0} \left(\frac{1}{2\pi} \int_{-\infty}^{\infty} \|f(\sigma+\imath\omega)\|_F^2 \,\mathrm{d}\omega \right)^{\frac{1}{2}} < \infty \right\}.$$

In particular, for a minimal standard LTI system (1.2), the associated *rational* transfer function

$$\mathbf{G}(\cdot) = \mathbf{C}(\cdot\mathbf{I} - \mathbf{A})^{-1}\mathbf{B} \in \mathcal{H}_2^{p\times m}$$

if and only if the eigenvalues of the system matrix $\mathbf{A}$ are located in the open left complex plane. In the context of model order reduction, the $\mathcal{H}_2$-norm is useful since it allows us to measure the approximation quality of a reduced-order model. Indeed, let us assume that y and y_r are the outputs obtained for (1.1) and (1.3) with zero initial condition. By (2.2), (2.3), Young's inequality for convolutions, and the Payley–Wiener theorem we then have

$$\begin{aligned} \|y - y_r\|_{L^\infty} &= \| \left(\mathbf{C}\Phi\mathbf{B} - \mathbf{C}_r\Phi_r\mathbf{B}_r\right) * u\|_{L^\infty} \leq \|\mathbf{C}\Phi\mathbf{B} - \mathbf{C}_r\Phi_r\mathbf{B}_r\|_{L^2}\|u\|_{L^2} \\ &= \|\mathbf{G} - \mathbf{G}_r\|_{\mathcal{H}_2^{p\times m}}\|u\|_{L^2}, \end{aligned} \tag{2.4}$$

where $*$ denotes the convolution of two functions. As it turns out, the approach discussed below provides an error bound for the reduced-order model with respect to the $\mathcal{H}_2$-norm.

2.3. Volterra equations as infinite-dimensional systems. Instead of (1.1), we may use a semigroup formulation and alternatively consider a Volterra integro-differential equation as an infinite-dimensional system. We briefly want to review the special case $\mu(\mathrm{d}s) = \mathbf{A}\delta_0(\mathrm{d}s) + \mathbf{K}(s)\,\mathrm{d}s$, where δ_0 is again a unit point mass at 0 and $\mathbf{K}(\cdot) \in L^1(0,\infty;\mathbb{R}^{n\times n})$. Hence, the state equation then reads

$$\begin{aligned} \dot{x}(t) &= \mathbf{A}x(t) + \int_0^t \mathbf{K}(s)x(t-s)\,\mathrm{d}s + \mathbf{B}u(t) \\ &= \mathbf{A}x(t) + \int_0^t \mathbf{K}(t-s)x(s)\,\mathrm{d}s + \mathbf{B}u(t). \end{aligned} \tag{2.5}$$

For what follows, we refer to the original presentation in [8]. For having a well-posed infinite-dimensional system, instead of (2.5) we rather consider the slightly modified version

$$\begin{aligned} \dot{x}(t) &= \mathbf{A}x(t) + \int_{-\infty}^t \mathbf{K}(t-s)x(s)\,\mathrm{d}s + \mathbf{B}u(t), \quad t \geq 0, \\ y(t) &= \mathbf{C}x(t). \end{aligned} \tag{2.6}$$

We now have to provide initial data

$$x(0) = \eta, \quad x(t) = \varphi(t), \quad t < 0.$$

As state space for system (2.6) we choose the product space

$$M^2 := \mathbb{R}^n \times L^2(-\infty, 0; \mathbb{R}^n).$$

We further define an operator $\mathcal{A}$ on M^2 by the domain

$$\mathcal{D}(\mathcal{A}) = \left\{ \begin{pmatrix} \eta \\ \varphi \end{pmatrix} \in M^2 \,|\varphi \in W^{1,2}(-\infty, 0; \mathbb{R}^n),\ \eta = \varphi(0) \right\}$$

and action

$$\mathcal{A} \begin{pmatrix} \eta \\ \varphi \end{pmatrix} = \begin{pmatrix} \mathbf{A}\eta + \int_{-\infty}^{0} \mathbf{K}(-s)\varphi(s)\, \mathrm{d}s \\ \varphi'(\cdot) \end{pmatrix}.$$

In particular, $\mathcal{A}$ is the infinitesimal generator of a C_0-semigroup $S(t)$ on M^2 [8, Theorem 2.1]. While this formulation can be carried out analogously for systems with finite delay, there occur some significant differences. For example, for systems with infinite delay, the spectrum is never discrete. In particular, we have that $\overline{\mathbb{C}}_- \subseteq \sigma(\mathcal{A})$ [8].

Remark 2.5. If in (2.5) it holds that $x(t) = 0$ for $t < 0$, the system can be seen as a special case of (2.6). However in the case of $x(0) \neq 0$ it follows that $(\eta, \phi) \notin \mathcal{D}(\mathcal{A})$. In other words we cannot expect to have classical solutions. Since for model reduction purposes one typically assumes that $x(0) = 0$, we will not tackle this situation in more detail.

If we define the control and observation operators by

$$\mathcal{B}\colon \mathbb{R}^m \to M^2, \quad \mathcal{B}u = \begin{pmatrix} \mathbf{B}u \\ 0 \end{pmatrix}, \quad \mathcal{C}\colon M^2 \to \mathbb{R}^p, \quad \mathcal{C} \begin{pmatrix} \eta \\ \varphi \end{pmatrix} = \mathbf{C}\eta$$

we can rewrite system (2.6) in the following form,

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t} z(t) &= \mathcal{A}z(t) + \mathcal{B}u(t), \quad z(0) = \begin{pmatrix} \eta \\ \varphi \end{pmatrix}, \\ y(t) &= \mathcal{C}z(t), \end{aligned} \tag{2.7}$$

where $z(t) = \begin{pmatrix} x(t) \\ x(t-\cdot) \end{pmatrix}$ denotes the state of the system at time t. From this point of view, $x(t)$ in (1.1) reflects only a part of the state and one might rather call it the position at time t; see also the discussion in [25].

3. Gramians and balancing transformations. For systems of the form (1.2), the method of balanced truncation constructs a reduced-order model by truncating a system that is balanced with respect to both its controllability and observability Gramians $\mathbf{X}$ and $\mathbf{Y}$, respectively. For an asymptotically stable system, it is well known [1] that $\mathbf{X}$ and $\mathbf{Y}$ are the unique positive-semidefinite solutions of the Lyapunov matrix equations

$$\mathbf{A}\mathbf{X} + \mathbf{X}\mathbf{A}^T + \mathbf{B}\mathbf{B}^T = 0, \quad \mathbf{A}^T\mathbf{Y} + \mathbf{Y}\mathbf{A} + \mathbf{C}^T\mathbf{C} = 0. \tag{3.1}$$

Moreover, alternative representations for $\mathbf{X}$ and $\mathbf{Y}$ are as follows:

$$\mathbf{X} = \int_0^\infty e^{\mathbf{A}t}\mathbf{B}\mathbf{B}^T e^{\mathbf{A}^T t}\,\mathrm{d}t = \frac{1}{2\pi}\int_{-\infty}^\infty (\imath\omega\mathbf{I}-\mathbf{A})^{-1}\mathbf{B}\mathbf{B}^T(-\imath\omega\mathbf{I}-\mathbf{A})^{-*}\mathrm{d}\omega,$$
$$\mathbf{Y} = \int_0^\infty e^{\mathbf{A}^T t}\mathbf{C}^T\mathbf{C}e^{\mathbf{A}t}\,\mathrm{d}t = \frac{1}{2\pi}\int_{-\infty}^\infty (\imath\omega\mathbf{I}-\mathbf{A})^{-*}\mathbf{C}^T\mathbf{C}(-\imath\omega\mathbf{I}-\mathbf{A})^{-1}\mathrm{d}\omega.$$

As a natural candidate for system (1.1), we now define the following Gramians

$$\mathbf{P} := \int_0^\infty \Phi(\tau)\mathbf{B}\mathbf{B}^T\Phi^T(\tau)\,\mathrm{d}\tau, \quad \mathbf{Q} := \int_0^\infty \Phi^T(\tau)\mathbf{C}^T\mathbf{C}\Phi(\tau)\,\mathrm{d}\tau, \tag{3.2}$$

where $\Phi(\cdot)$ denotes the differential resolvent. As we pointed out in section 2, if system (1.1) is stable, we have that $\Phi(\cdot) \in L^2(0,\infty;\mathbb{R}^{n\times n})$. Consequently, the expressions in (3.2) are indeed well defined since, e.g.,

$$\|\mathbf{P}\|_2 = \left\|\int_0^\infty \Phi(\tau)\mathbf{B}\mathbf{B}^T\Phi^T(\tau)\,\mathrm{d}\tau\right\|_2 \le \|\mathbf{B}\|_2^2\int_0^\infty \|\Phi(\tau)\|_2^2\,\mathrm{d}\tau < \infty.$$

Note that the definitions of $\mathbf{P}$ and $\mathbf{Q}$ are straightforward extensions of the Gramians defined for finite delay systems in [25]. We want to characterize $\mathbf{P}$ and $\mathbf{Q}$ in terms of (generalized) matrix equations analogously to (3.1). Following similar works on finite delay systems [26, 27, 28, 37], we therefore consider $\mathbf{P}$ and $\mathbf{Q}$ as special cases arising from the *delay Lyapunov matrices*

$$\mathbf{P}(t) := \int_0^\infty \Phi(\tau)\mathbf{B}\mathbf{B}^T\Phi^T(\tau+t)\,\mathrm{d}\tau, \quad \mathbf{Q}(t) := \int_0^\infty \Phi^T(\tau)\mathbf{C}^T\mathbf{C}\Phi(\tau+t)\,\mathrm{d}\tau. \tag{3.3}$$

In [37, Proposition 6.28], for the finite delay case these matrices have been shown to satisfy a matrix delay differential equation together with a symmetry and an algebraic condition, respectively. As the next proposition shows, for our generalized Volterra integro-differential setting, the above matrices still satisfy a matrix delay differential equation which now is of Volterra type. Moreover, the symmetry condition as well as the algebraic condition are retained; the latter, however, with a slight extension to the integro-differential character considered here.

PROPOSITION 3.1. *Let* $\mu \in \mathcal{M}(0,\infty;\mathbb{R}^{n\times n})$. *Assume that* $\det(z\mathbf{I}-\hat{\mu}(z)) \neq 0$, $\mathrm{Re}(z) \ge 0$. *Then* $\mathbf{P}(\cdot), \mathbf{Q}(\cdot)$ *almost everywhere satisfy the matrix delay differential equations*

$$\begin{aligned}\dot{\mathbf{P}}(t) &= \int_0^\infty \mathbf{P}(t-s)\mu^T(\mathrm{d}s), \quad t \ge 0,\\ \dot{\mathbf{Q}}(t) &= \int_0^\infty \mathbf{Q}(t-s)\mu(\mathrm{d}s), \quad t \ge 0,\end{aligned} \tag{3.4}$$

$$\mathbf{P}(t) = \mathbf{P}^T(-t), \quad \mathbf{Q}(t) = \mathbf{Q}^T(-t), \quad t \ge 0, \tag{3.5}$$

$$\begin{aligned}0 &= \int_0^\infty \mathbf{P}^T(s)\mu^T(\mathrm{d}s) + \int_0^\infty \mu(\mathrm{d}s)\mathbf{P}(s) + \mathbf{B}\mathbf{B}^T,\\ 0 &= \int_0^\infty \mathbf{Q}^T(s)\mu(\mathrm{d}s) + \int_0^\infty \mu^T(\mathrm{d}s)\mathbf{Q}(s) + \mathbf{C}^T\mathbf{C}.\end{aligned} \tag{3.6}$$

Proof. We only show the assertion for $\mathbf{P}$. Analogous reasoning can be used to prove the equation for $\mathbf{Q}$. We recall that $\Phi(t) = 0$ for $t < 0$. A simple calculation then leads to

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{P}(t) &= \frac{\mathrm{d}}{\mathrm{d}t}\int_0^\infty \Phi(\tau)\mathbf{B}\mathbf{B}^T\Phi^T(\tau+t)\,\mathrm{d}\tau \\ &= \int_0^\infty \Phi(\tau)\mathbf{B}\mathbf{B}^T\left(\int_0^{\tau+t}\mu(\mathrm{d}s)\Phi(\tau+t-s)\right)^T\mathrm{d}\tau \\ &= \int_0^\infty \Phi(\tau)\mathbf{B}\mathbf{B}^T\left(\int_0^{\tau+t}\Phi^T(\tau+t-s)\mu^T(\mathrm{d}s)\right)\mathrm{d}\tau.\end{aligned}$$

Changing the order of integration yields

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{P}(t) &= \int_0^t\int_0^\infty \Phi(\tau)\mathbf{B}\mathbf{B}^T\Phi^T(\tau+t-s)\,\mathrm{d}\tau\,\mu^T(\mathrm{d}s) \\ &\quad + \int_t^\infty\int_{s-t}^\infty \Phi(\tau)\mathbf{B}\mathbf{B}^T\Phi^T(\tau+t-s)\,\mathrm{d}\tau\,\mu^T(\mathrm{d}s) \\ &= \int_0^t \mathbf{P}(t-s)\mu^T(\mathrm{d}s) + \int_t^\infty\int_{s-t}^\infty \Phi(\tau)\mathbf{B}\mathbf{B}^T\Phi^T(\tau+t-s)\,\mathrm{d}\tau\,\mu^T(\mathrm{d}s).\end{aligned}$$

Since $\Phi(t) = 0, t < 0$, this is the same as

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{P}(t) &= \int_0^t \mathbf{P}(t-s)\mu^T(\mathrm{d}s) + \int_t^\infty\int_0^\infty \Phi(\tau)\mathbf{B}\mathbf{B}^T\Phi^T(\tau+t-s)\,\mathrm{d}\tau\,\mu^T(\mathrm{d}s) \\ &= \int_0^\infty \mathbf{P}(t-s)\mu^T(\mathrm{d}s).\end{aligned}$$

This shows (3.4). For (3.5), note that due to (3.3), we have that

$$\begin{aligned}\mathbf{P}(-t) &= \int_0^\infty \Phi(\tau)\mathbf{B}\mathbf{B}^T\Phi^T(\tau-t)\,\mathrm{d}\tau = \int_{-t}^\infty \Phi(s+t)\mathbf{B}\mathbf{B}^T\Phi^T(s)\,\mathrm{d}s \\ &= \int_0^\infty \Phi(s+t)\mathbf{B}\mathbf{B}^T\Phi^T(s)\,\mathrm{d}s = \mathbf{P}(t)^T.\end{aligned}$$

Finally, we find that

$$\begin{aligned}&\int_0^\infty \mathbf{P}^T(s)\mu^T(\mathrm{d}s) + \int_0^\infty \mu(\mathrm{d}s)\mathbf{P}(s) = \int_0^\infty \mathbf{P}(-s)\mu^T(\mathrm{d}s) + \int_0^\infty \mu(\mathrm{d}s)\mathbf{P}^T(-s) \\ &\quad = \int_0^\infty\left(\int_0^\infty \Phi(\tau)\mathbf{B}\mathbf{B}^T\Phi(\tau-s)^T\,\mathrm{d}\tau\right)\mu^T(\mathrm{d}s) \\ &\qquad + \int_0^\infty \mu(\mathrm{d}s)\left(\int_0^\infty \Phi^T(\tau-s)\mathbf{B}\mathbf{B}^T\Phi(\tau)\,\mathrm{d}\tau\right) \\ &\quad = \int_0^\infty\left(\int_s^\infty \Phi(\tau)\mathbf{B}\mathbf{B}^T\Phi^T(\tau-s)\,\mathrm{d}\tau\right)\mu^T(\mathrm{d}s) \\ &\qquad + \int_0^\infty \mu(\mathrm{d}s)\left(\int_s^\infty \Phi^T(\tau-s)\mathbf{B}\mathbf{B}^T\Phi(\tau)\,\mathrm{d}\tau\right)\end{aligned}$$

$$
\begin{aligned}
&= \int_0^\infty \left(\int_0^\tau \Phi(\tau)\mathbf{B}\mathbf{B}^T\Phi^T(\tau - s)\mu^T(\mathrm{d}s) \right) \mathrm{d}\tau \\
&\quad + \int_0^\infty \left(\int_0^\tau \mu(\mathrm{d}s)\Phi^T(\tau - s)\mathbf{B}\mathbf{B}^T\Phi(\tau)^T \right) \mathrm{d}\tau \\
&= \int_0^\infty \frac{\mathrm{d}}{\mathrm{d}\tau}(\Phi(\tau)\mathbf{B}\mathbf{B}^T\Phi^T(\tau))\,\mathrm{d}\tau = -\mathbf{B}\mathbf{B}^T. \qquad \square
\end{aligned}
$$

Remark 3.2. Note that the results from Proposition 3.1 are indeed consistent with existing results from the literature. For example, in the case of a finite delay system, we have that $\mu(\mathrm{d}s) = \delta_0(\mathrm{d}s)\mathbf{A} + \delta_\tau(\mathrm{d}s)\mathbf{A}_1$. Hence, it follows that

$$
\begin{aligned}
\dot{\mathbf{P}}(t) &= \int_0^\infty \mathbf{P}(t-s)\mu^T(\mathrm{d}s) = \mathbf{P}(t)\mathbf{A}^T + \mathbf{P}(t-\tau)\mathbf{A}_1^T, \\
0 &= \int_0^\infty \mathbf{P}^T(s)\mu^T(\mathrm{d}s) + \int_0^\infty \mu(\mathrm{d}s)\mathbf{P}(s) + \mathbf{B}\mathbf{B}^T \\
&= \mathbf{P}(0)\mathbf{A}^T + \mathbf{P}(-\tau)\mathbf{A}_1^T + \mathbf{A}\mathbf{P}(0) + \mathbf{A}_1\mathbf{P}(\tau) + \mathbf{B}\mathbf{B}^T.
\end{aligned}
$$

These delay Lyapunov equations are exactly those that have been discussed in, e.g., [25, 37]. Equation (3.4) has been studied even earlier in the context of stability theory for difference-differential equations; see [10].

Remark 3.3. Let us point out that the differential equations (3.4) for $\mathbf{P}(t)$ and $\mathbf{Q}(t)$ only depend on previous time instances. In fact, as in (2.6) consider again the case where $\mu(\mathrm{d}s) = \mathbf{A}\delta_0(\mathrm{d}s) + \mathbf{K}(s)\mathrm{d}s$. The differential equation for $\mathbf{P}(t)$ then reads

$$
\dot{\mathbf{P}}(t) = \mathbf{P}(t)\mathbf{A}^T + \int_0^\infty \mathbf{P}(t-s)\mathbf{K}^T(s)\,\mathrm{d}s.
$$

A simple change of variables implies that

$$
\dot{\mathbf{P}}(t) = \mathbf{P}(t)\mathbf{A}^T + \int_{-\infty}^t \mathbf{P}(t)\mathbf{K}^T(t-s)\,\mathrm{d}s
$$

which shows the similarity to the underlying system (2.6).

We further have a characterization of $\mathbf{P}(\cdot)$ and $\mathbf{Q}(\cdot)$ in terms of an integral along the imaginary axis. Again, the result can be found for finite delay systems in [37, Proposition 6.29].

PROPOSITION 3.4. *Let* $\mu \in \mathcal{M}(0,\infty;\mathbb{R}^{n\times n})$. *Assume that* $\det(z\mathbf{I} - \hat{\mu}(z)) \neq 0$, $\mathrm{Re}(z) \geq 0$. *For* $z \in \overline{\mathbb{C}}_+$, *define* $\mathbf{R}(z) := (z\mathbf{I} - \hat{\mu}(z))^{-1}$. *Then*

$$
\begin{aligned}
\mathbf{P}(t) &= \frac{1}{2\pi}\int_{-\infty}^\infty \mathbf{R}(\imath\omega)\mathbf{B}\mathbf{B}^T\mathbf{R}^*(\imath\omega)e^{\imath\omega t}\,\mathrm{d}\omega, \\
\mathbf{Q}(t) &= \frac{1}{2\pi}\int_{-\infty}^\infty \mathbf{R}^*(\imath\omega)\mathbf{C}^T\mathbf{C}\mathbf{R}(\imath\omega)e^{\imath\omega t}\,\mathrm{d}\omega
\end{aligned} \tag{3.7}
$$

for all $t \in \mathbb{R}$.

Proof. By Theorem 2.3 and [19, Theorem 3.3.9] we have $\Phi(\cdot) \in L^1(0,\infty;\mathbb{R}^{n\times n}) \cap L^2(0,\infty;\mathbb{R}^{n\times n})$. Hence, the Fourier transform of the differential resolvent $\hat{\Phi}(\omega) = \int_0^\infty \Phi(t)e^{-\imath\omega t}\,\mathrm{d}t$ exists. In fact, since $\mu \in \mathcal{M}(0,\infty;\mathbb{R}^{n\times n})$, applying the Fourier transformation to (1.1) shows that $\hat{\Phi}(\omega) = \mathbf{R}(\imath\omega)$. Consequently, the integrals in (3.7)

are well defined. The assertion now follows by the arguments provided [37, Proposition 6.29] for the finite delay case. For arbitrary $x, y \in \mathbb{C}^n$ using Parseval's identity (e.g., [19, Theorem 16.8.2]) we obtain that

$$\begin{aligned} y^* \mathbf{P}(t) x &= y^* \left(\int_0^\infty \Phi(\tau) \mathbf{B}\mathbf{B}^T \Phi^T(\tau + t) \, \mathrm{d}\tau \right) x \\ &= \int_0^\infty \left\langle \mathbf{B}^T \Phi^T(\tau + t) x, \mathbf{B}^T \Phi^T(\tau) y \right\rangle_{\mathbb{C}^n} \, \mathrm{d}\tau. \end{aligned}$$

Using that $\Phi(-t) = 0, t > 0$, we continue with

$$\begin{aligned} &= \int_{-\infty}^\infty \left\langle \mathbf{B}^T \Phi^T(\tau + t) x, \mathbf{B}^T \Phi^T(\tau) y \right\rangle_{\mathbb{C}^n} \, \mathrm{d}\tau \\ &= \frac{1}{2\pi} \int_{-\infty}^\infty \left\langle \mathbf{B}^T \mathbf{R}^*(\imath\omega) x, \mathbf{B}^T \mathbf{R}^*(\imath\omega) y \right\rangle_{\mathbb{C}^n} e^{\imath\omega t} \, \mathrm{d}\omega \\ &= y^* \left(\frac{1}{2\pi} \int_{-\infty}^\infty \mathbf{R}(\imath\omega) \mathbf{B}\mathbf{B}^T \mathbf{R}(\imath\omega)^* e^{\imath\omega t} \, \mathrm{d}\omega \right) x. \end{aligned}$$

This shows the assertion for $\mathbf{P}(\cdot)$. The same arguments can be used for $\mathbf{Q}(\cdot)$. ∎

Remark 3.5. Note that Proposition 3.4 assumes that $\mu \in \mathcal{M}(0, \infty; \mathbb{R}^{n \times n})$. As we will see below, some applications involve an unbounded measure $\mu_{\mathrm{loc}}(0, \infty; \mathbb{R}^{n \times n})$. Under the assumptions in Theorem 2.4, we still have that $\Phi(\cdot) \in L^2(0, \infty; \mathbb{R}^{n \times n})$ such that its Fourier transform can be defined. Unfortunately, the meaning of $\hat{\mu}(\imath\omega)$ is not clear a priori. We however comment on that later on in more detail.

Since for typical balancing based model reduction methods we are interested in an energy interpretation of the Gramians, it is useful to associate $\mathbf{P}$ and $\mathbf{Q}$ with their finite horizon counterparts

$$\mathbf{P}_{t_f} := \int_0^{t_f} \Phi(\tau) \mathbf{B}\mathbf{B}^T \Phi^T(\tau) \, \mathrm{d}\tau, \quad \mathbf{Q}_{t_f} := \int_0^{t_f} \Phi(\tau)^T \mathbf{C}^T \mathbf{C} \Phi(\tau) \, \mathrm{d}\tau. \tag{3.8}$$

Moreover, given a system of the form (1.1) with initial data $x(0) = x_0$ and input function u we denote by $x(t, x_0, u)$ and $y(t, x_0, 0)$ the solution and the output of the system at time t. Let $x_1 \in \mathbb{R}^n$ denote a given state that should be reached. The associated *energies* E_c and E_o are further defined as follows,

$$\begin{aligned} E_c^{t_f}(x_1) &:= \min_{\substack{u \in L^2(0, t_f; \mathbb{R}^m) \\ x(t_f, 0, u) = x_1}} \int_0^{t_f} \|u(t)\|^2 \, \mathrm{d}t, \\ E_o^{t_f}(x_0) &:= \int_0^{t_f} \|y(t, x_0, 0)\|^2 \, \mathrm{d}t. \end{aligned} \tag{3.9}$$

A connection between the energies in (3.9) and the Gramians in (3.8) is as follows.

THEOREM 3.6. *Let $x_0 \in \mathbb{R}^n$ and $x_1 \in \mathrm{Im}(\mathbf{P}_{t_f})$ be given. Define $u_{\mathrm{opt}} \colon [0, t_f] \to \mathbb{R}^m$ by $u_{\mathrm{opt}}(t) = \mathbf{B}^T \Phi(t_f - t)^T \mathbf{P}_{t_f}^\dagger x_1$. Let $\tilde{u} \in L^2(0, t_f; \mathbb{R}^m)$ be an arbitrary control steering the system from the zero state to x_1 in time t_f. Then*

$$\begin{aligned} E_c^{t_f}(x_1) &= \|u_{\mathrm{opt}}\|_{L^2}^2 = \langle x_1, \mathbf{P}_{t_f}^\dagger x_1 \rangle \le \|\tilde{u}\|_{L^2}^2, \\ E_o^{t_f}(x_0) &= \langle x_0, \mathbf{Q}_{t_f} x_0 \rangle. \end{aligned}$$

Proof. Using the variation of constants formula (2.2), for $x_1 \in \mathrm{Im}(\mathbf{P}_{t_f})$ it holds that

$$\begin{aligned} x(t_f, 0, u_{\mathrm{opt}}) &= \int_0^{t_f} \Phi(t_f - s)\mathbf{B}u_{\mathrm{opt}}(s)\,\mathrm{d}s \\ &= \int_0^{t_f} \Phi(t_f - s)\mathbf{B}\mathbf{B}^T\Phi(t_f - s)^T\mathbf{P}_{t_f}^{\dagger}x_1\,\mathrm{d}s \\ &= \mathbf{P}_{t_f}\mathbf{P}_{t_f}^{\dagger}x_1 = x_1. \end{aligned}$$

We further observe that

$$\begin{aligned} \|u_{\mathrm{opt}}\|_{L^2}^2 &= \int_0^{t_f} x_1^T\mathbf{P}_{t_f}^{\dagger}\Phi(t_f - t)\mathbf{B}\mathbf{B}^T\Phi(t_f - t)^T\mathbf{P}_{t_f}^{\dagger}x_1\,\mathrm{d}t \\ &= x_1^T\mathbf{P}_{t_f}^{\dagger}\int_0^{t_f}\Phi(t_f - t)\mathbf{B}\mathbf{B}^T\Phi(t_f - t)^T\,\mathrm{d}t\,\mathbf{P}_{t_f}^{\dagger}x_1 \\ &= x_1^T\mathbf{P}_{t_f}^{\dagger}\mathbf{P}_{t_f}\mathbf{P}_{t_f}^{\dagger}x_1 = \langle x_1, \mathbf{P}_{t_f}^{\dagger}x_1\rangle. \end{aligned}$$

Assume now that $\tilde{u}$ is another control with $x(t_f, 0, \tilde{u}) = x_1$. We then have

$$0 = x(t_f, 0, \tilde{u}) - x(t_f, 0, u_{\mathrm{opt}}) = \int_0^{t_f}\Phi(t_f - t)\mathbf{B}(\tilde{u}(t) - u_{\mathrm{opt}}(t))\,\mathrm{d}t.$$

Multiplication with $x_1^T\mathbf{P}_{t_f}^{\dagger}$ from the left yields $\langle u_{\mathrm{opt}}, \tilde{u} - u_{\mathrm{opt}}\rangle_{L^2} = 0$. Hence, we conclude that

$$\begin{aligned} \langle\tilde{u}, \tilde{u}\rangle &= \langle(\tilde{u} - u_{\mathrm{opt}}) + u_{\mathrm{opt}}, (\tilde{u} - u_{\mathrm{opt}}) + u_{\mathrm{opt}}\rangle \\ &= \langle\tilde{u} - u_{\mathrm{opt}}, \tilde{u} - u_{\mathrm{opt}}\rangle + \langle u_{\mathrm{opt}}, u_{\mathrm{opt}}\rangle \geq \langle u_{\mathrm{opt}}, u_{\mathrm{opt}}\rangle. \end{aligned}$$

This show the first assertion. The second statement also follows by (2.2). Let $x_0 \in \mathbb{R}^n$ be given. We then have

$$\begin{aligned} E_o^{t_f}(x_0) &= \int_0^{t_f}\|y(t, x_0, 0)\|^2\,\mathrm{d}t = \int_0^{t_f}\|\mathbf{C}x(t, x_0, 0)\|^2\,\mathrm{d}t \\ &= \int_0^{t_f}\|\mathbf{C}\Phi(t)x_0\|^2\,\mathrm{d}t = \langle x_0, \mathbf{Q}_{t_f}x_0\rangle. \end{aligned}$$ ∎

As in [25, Theorem 1], we have the following asymptotic result.

COROLLARY 3.7. *Consider a stable system of the form* (1.1). *For* $t_f \to \infty$, *the energies in* (3.9) *are given by*

$$E_c^{\infty}(x_1) = \begin{cases} \langle x_1, \mathbf{P}^{\dagger}x_1\rangle & \text{if } x_1 \in \mathrm{Im}(\mathbf{P}), \\ \infty & \text{if } x_1 \notin \mathrm{Im}(\mathbf{P}), \end{cases} \qquad E_o^{\infty}(x_0) = \langle x_0, \mathbf{Q}x_0\rangle.$$

Remark 3.8. In analogy to [25, Remark 2], we mention that from an infinite-dimensional point of view, the energy concepts (3.9) have to be slightly modified. Since the state of the infinite-dimensional system is $z(t) = \begin{pmatrix} x(t) \\ x(t-\cdot)\end{pmatrix}$, we obtain the free history optimization problem

$$E_c^{t_f}(x_1) := \inf_{\substack{\varphi\in L^2(-\infty,0;\mathbb{R}^n) \\ x(t_f+\tau)=\varphi(\tau),\tau\in(-\infty,0)}} \inf_{\substack{u\in L^2(0,t_f;\mathbb{R}^m) \\ x(t_f,0,u)=x_1}} \int_0^{t_f}\|u(t)\|^2\,\mathrm{d}t. \tag{3.10}$$

Similarly, in addition to $x(0) = x_0$, for the output energy we assume that $x(t) = 0, t < 0$.

We can now apply the concept of balancing to equations of the form (1.1). As for linear systems the idea is to find a state space transformation of (1.1) such that

$$\mathbf{P} = \mathbf{Q} = \mathbf{\Sigma} = \operatorname{diag}(\sigma_1, \dots, \sigma_n).$$

In this way, we expect that states associated with "small" values σ_i can be neglected without significantly changing the input-output behavior of the original system. In other words, given a balanced system, a reduced-order model of the form (1.3) is obtained by simple truncation. Assume that system (1.1) is subject to a state space transformation

$$(\mu, \mathbf{B}, \mathbf{C}) \to (\mathbf{T}\mu\mathbf{T}^{-1}, \mathbf{TB}, \mathbf{CT}^{-1}).$$

Using Proposition 3.4, for the Gramians $\tilde{\mathbf{P}}$ and $\tilde{\mathbf{Q}}$ of the transformed system it holds that

$$\tilde{\mathbf{P}} = \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathbf{TR}(\imath\omega)\mathbf{T}^{-1}\mathbf{TBB}^T\mathbf{T}^T\mathbf{T}^{-T}\mathbf{R}^*(\imath\omega)\mathbf{T}^T \mathrm{d}\omega = \mathbf{TPT}^T,$$

$$\tilde{\mathbf{Q}} = \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathbf{T}^{-T}\mathbf{R}^*(\imath\omega)\mathbf{T}^T\mathbf{T}^{-T}\mathbf{C}^T\mathbf{CT}^{-1}\mathbf{TR}(\imath\omega)\mathbf{T}^{-1} \mathrm{d}\omega = \mathbf{T}^{-T}\mathbf{QT}^{-1}.$$

In other words, the Gramians are subject to a *contragredient transformation*. In this case, the well-known scheme for obtaining a balanced realization is as follows. Assume that the Cholesky decompositions of $\mathbf{P}$ and $\mathbf{Q}$ are given:

$$\mathbf{P} = \mathbf{L}_P\mathbf{L}_P^T, \quad \mathbf{Q} = \mathbf{L}_Q\mathbf{L}_Q^T, \quad \mathbf{L}_P, \mathbf{L}_Q \in \mathbb{R}^{n\times n}.$$

Computing a singular value decomposition of $\mathbf{L}_P^T\mathbf{L}_Q$ yields

$$\mathbf{U\Sigma V}^T = \mathbf{L}_P^T\mathbf{L}_Q$$

with orthonormal matrices $\mathbf{U}, \mathbf{V} \in \mathbb{R}^{n\times n}$ and diagonal matrix $\mathbf{\Sigma} \in \mathbb{R}^{n\times n}$. Hence, if we choose $\mathbf{T} = \mathbf{\Sigma}^{-\frac{1}{2}}\mathbf{V}^T\mathbf{L}_Q^T$ and note that $\mathbf{T}^{-1} = \mathbf{L}_P\mathbf{U}\mathbf{\Sigma}^{-\frac{1}{2}}$ for the transformed system, we find that $\tilde{\mathbf{P}} = \mathbf{\Sigma} = \tilde{\mathbf{Q}}$. A reduced-order model then can be obtained by simple truncation. We summarize the required steps in Algorithm 1.

Remark 3.9. In [25], the authors have provided a relation between $\mathbf{P}$ and $\mathbf{Q}$ and the Gramians $\mathcal{P}$ and $\mathcal{Q}$ of the infinite-dimensional system (2.7). A generalization to the setting considered here seems problematic due to the spectral properties of $\mathcal{A}$ in (2.7). Without fading memory assumptions on the measure μ, system (2.7) is, in general, not exponentially stable and the existence of $\mathcal{P}$ and $\mathcal{Q}$ is not guaranteed.

Algorithm 1. Balanced truncation for Volterra integro-differential equations.

Input: $\mu, \mathbf{B}, \mathbf{C}$ as in (1.1) such that Theorem 2.4 holds
Output: $\mu_r, \mathbf{B}_r, \mathbf{C}_r$

1: Compute $\mathbf{P}$ and $\mathbf{Q}$
2: Compute Cholesky decompositions $\mathbf{P} = \mathbf{L}_P\mathbf{L}_P^T, \quad \mathbf{Q} = \mathbf{L}_Q\mathbf{L}_Q^T$
3: Compute singular value decomposition $\mathbf{U\Sigma V}^T = \mathbf{L}_P^T\mathbf{L}_Q$
4: Set $\mathbf{W}_r = \mathbf{L}_Q\mathbf{V}_{(:,1:r)}\mathbf{\Sigma}_{(1:r,1:r)}^{-\frac{1}{2}}$
5: Set $\mathbf{V}_r = \mathbf{L}_P\mathbf{U}_{(:,1:r)}\mathbf{\Sigma}_{(1:r,1:r)}^{-\frac{1}{2}}$
6: Set $\mu_r = \mathbf{W}_r^T\mu\mathbf{V}_r, \quad \mathbf{B}_r = \mathbf{W}_r^T\mathbf{B}, \quad \mathbf{C}_r = \mathbf{CV}_r$

An error bound for the $\mathcal{H}_2$-norm. With regard to (2.4), let us analyze the properties of a reduced-order model resulting from Algorithm 1 in more detail. For this, we follow the strategy used in [42]. While the latter work focuses on structured systems of a certain type, e.g., weighted, second order, closed loop, we can extend the idea to our setting as well. First, we note that the reduced-order model μ_r actually is obtained by truncating the *balanced* system $(\mu_{\mathcal{B}}, \mathbf{B}_{\mathcal{B}}, \mathbf{C}_{\mathcal{B}})$. We thus consider the following partitioning,

$$\mu_{\mathcal{B}} = \begin{pmatrix} \mu_r & \mu_{12} \\ \mu_{21} & \mu_{22} \end{pmatrix}, \quad \mathbf{B}_{\mathcal{B}} = \begin{pmatrix} \mathbf{B}_r \\ \mathbf{B}_2 \end{pmatrix}, \quad \mathbf{C}_{\mathcal{B}} = \begin{pmatrix} \mathbf{C}_r & \mathbf{C}_2 \end{pmatrix}. \tag{3.11}$$

Further let us denote the *error system* by

$$\mu_{\mathcal{E}} = \begin{pmatrix} \mu_{\mathcal{B}} & 0 \\ 0 & \mu_r \end{pmatrix}, \quad \mathbf{B}_{\mathcal{E}} = \begin{pmatrix} \mathbf{B}_{\mathcal{B}} \\ \mathbf{B}_r \end{pmatrix}, \quad \mathbf{C}_{\mathcal{E}} = \begin{pmatrix} \mathbf{C}_{\mathcal{B}} & -\mathbf{C}_r \end{pmatrix}. \tag{3.12}$$

With this notation, we are interested in a bound for

$$\|\mathbf{G} - \mathbf{G}_r\|_{\mathcal{H}_2} = \|\mathbf{G}_{\mathcal{E}}\|_{\mathcal{H}_2} = \left\|\mathbf{C}_{\mathcal{E}}(\cdot\mathbf{I} - \hat{\mu}_{\mathcal{E}}(\cdot))^{-1}\mathbf{B}_{\mathcal{E}}\right\|_{\mathcal{H}_2} = \|\mathbf{C}_{\mathcal{E}}\mathbf{R}_{\mathcal{E}}(\cdot)\mathbf{B}_{\mathcal{E}}\|_{\mathcal{H}_2}.$$

Let us set $\mathbf{N}_{\mathcal{B}}(\imath\omega) = \left(\begin{smallmatrix} \mathbf{N}_{\mathcal{B},1}(\imath\omega) \\ \mathbf{N}_{\mathcal{B},2}(\imath\omega) \end{smallmatrix}\right) := \mathbf{R}_{\mathcal{B}}(\imath\omega)\mathbf{B}_{\mathcal{B}}$ and $\mathbf{N}_r(\imath\omega) := \mathbf{R}_r(\imath\omega)\mathbf{B}_r$. For the balanced system, we now conclude that

$$\begin{aligned} \mathbf{\Sigma}_{(1:r,1:r)} &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathbf{N}_{\mathcal{B},1}(\imath\omega)\mathbf{N}_{\mathcal{B},1}(\imath\omega)^* \,\mathrm{d}\omega, \\ \mathbf{\Sigma}_{(r+1:n,r+1:n)} &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathbf{N}_{\mathcal{B},2}(\imath\omega)\mathbf{N}_{\mathcal{B},2}(\imath\omega)^* \,\mathrm{d}\omega, \\ 0 &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathbf{N}_{\mathcal{B},1}(\imath\omega)\mathbf{N}_{\mathcal{B},2}(\imath\omega)^* \,\mathrm{d}\omega. \end{aligned}$$

Moreover, we find $\mathbf{R}_{\mathcal{B}}^{-1}(\imath\omega)\mathbf{N}_{\mathcal{B}}(\imath\omega) = \mathbf{B}_{\mathcal{B}}$ which implies that

$$(\imath\omega\mathbf{I} - \hat{\mu}_r(\imath\omega))\mathbf{N}_{\mathcal{B},1}(\imath\omega) - \hat{\mu}_{12}(\imath\omega)\mathbf{N}_{\mathcal{B},2}(\imath\omega) = \mathbf{B}_r$$

and, consequently,

$$\mathbf{N}_r(\imath\omega) = (\imath\omega\mathbf{I} - \hat{\mu}_r(\imath\omega))^{-1}\mathbf{B}_r = \mathbf{N}_{\mathcal{B},1}(\imath\omega) - \underbrace{(\imath\omega\mathbf{I} - \hat{\mu}_r(\imath\omega))^{-1}\hat{\mu}_{12}(\imath\omega)}_{:=\mathbf{H}(\imath\omega)}\mathbf{N}_{\mathcal{B},2}(\imath\omega).$$

Due to the definition of the $\mathcal{H}_2$-norm, it then holds

$$\begin{aligned} \|\mathbf{G}_{\mathcal{E}}\|_{\mathcal{H}_2}^2 = \frac{1}{2\pi}\mathrm{tr}\left(\int_{-\infty}^{\infty} \mathbf{C}_{\mathcal{B}}\mathbf{N}_{\mathcal{B}}(\imath\omega)\mathbf{N}_{\mathcal{B}}(\imath\omega)^*\mathbf{C}_{\mathcal{B}}^T \,\mathrm{d}\omega\right) \\ - 2\frac{1}{2\pi}\mathrm{tr}\left(\int_{-\infty}^{\infty} \mathbf{C}_{\mathcal{B}}\mathbf{N}_{\mathcal{B}}(\imath\omega)\mathbf{N}_r(\imath\omega)^*\mathbf{C}_r^T \,\mathrm{d}\omega\right) \\ + \frac{1}{2\pi}\mathrm{tr}\left(\int_{-\infty}^{\infty} \mathbf{C}_r\mathbf{N}_r(\imath\omega)\mathbf{N}_r(\imath\omega)^*\mathbf{C}_r^T \,\mathrm{d}\omega\right). \end{aligned}$$

As shown similarly in [42], the above properties of the balanced system lead to the simplified expression

$$\begin{aligned} \|\mathbf{G}_{\mathcal{E}}\|_{\mathcal{H}_2}^2 = \mathrm{tr}\left(\mathbf{C}_2\mathbf{\Sigma}_{(r+1:n,r+1:n)}\mathbf{C}_2^T\right) \\ + \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{tr}\left((\mathbf{C}_r\mathbf{H}(\imath\omega) - 2\mathbf{C}_2)\mathbf{N}_{\mathcal{B},2}(\imath\omega)(\mathbf{C}_r\mathbf{H}(\imath\omega)\mathbf{N}_{\mathcal{B},2}(\imath\omega))^* \,\mathrm{d}\omega\right). \end{aligned}$$

Finally, we may estimate this term in order to obtain

$$\begin{aligned}\|\mathbf{G}_{\mathcal{E}}\|_{\mathcal{H}_2}^2 &\leq \operatorname{tr}\left(\mathbf{C}_2\boldsymbol{\Sigma}_{(r+1:n,r+1:n)}\mathbf{C}_2^T\right)\\ &\quad + \sup_{\omega}\|\mathbf{H}(\imath\omega)^*\mathbf{C}_r^*(\mathbf{C}_r\mathbf{H}(\imath\omega)-2\mathbf{C}_2)\|_2\operatorname{tr}\left(\boldsymbol{\Sigma}_{(r+1:n,r+1:n)}\right)\end{aligned} \tag{3.13}$$

which has the same form as the $\mathcal{H}_2$-error bound for structured systems derived in [42].

Some remarks concerning the error bound are in order. Note that (3.13) relates the $\mathcal{H}_2$-norm of the error system with the neglected part of the (balanced) system Gramian. Hence, if the eigenvalues of $\boldsymbol{\Sigma}$ exhibit a *fast* decay, we expect the reduced-order system to be a good approximation of the original system in terms of the $\mathcal{H}_2$-norm. Moreover, let us emphasize that, once the balanced system is obtained, the complexity of computing (3.13) is essentially depending on the reduced-order dimension r. Indeed, computing the spectral norm can be efficiently done due to the decomposition into factors of rank at most r. Finally, we point out that a (more common) $\mathcal{H}_\infty$-error bound is not readily available due to the structure-preserving reduction technique.

4. Special choices of $\boldsymbol{\mu}$. In this section, we first consider two examples that are concerned with coupled as well as time fractional control systems. The other examples show that our results are indeed consistent with existing results from the literature.

4.1. Kernels of exponential type. Consider a measure μ given as the sum of a point mass at 0 and an exponential Volterra kernel

$$\mu(\mathrm{d}s) = \mathbf{A}\delta_0(\mathrm{d}s) + \sum_{i=1}^{k} e^{-\gamma_i s}\mathbf{K}_i\ \mathrm{d}s, \gamma_i > 0.$$

For simplicity, assume that $k = 1$. As a consequence, (1.1) reads as follows

$$\begin{aligned}\dot{x}(t) &= \mathbf{A}x(t) + \int_0^t e^{-\gamma_1(t-s)}\mathbf{K}_1 x(s)\ \mathrm{d}s + \mathbf{B}u(t),\\ y(t) &= \mathbf{C}x(t), \quad x(0) = x_0.\end{aligned} \tag{4.1}$$

Typical examples leading to a system of the form (4.1) are the discretized linear heat equation with fading memory [31, 35] as well as linearized versions of the FitzHugh–Nagumo system [30] frequently used in cardiac electrophsyiology.

Introduction of a new variable $z(t) := \int_0^t e^{-\gamma_1(t-s)}\mathbf{K}_1 x(s)\ \mathrm{d}s$ and differentiation also yield

$$\dot{z}(t) = \mathbf{K}_1 x(t) - \gamma_1 z(t), \quad z(0) = 0.$$

Hence, instead of (4.1) we may consider the coupled system

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix}x(t)\\ z(t)\end{pmatrix} &= \underbrace{\begin{pmatrix}\mathbf{A} & \mathbf{I}\\ \mathbf{K}_1 & -\gamma_1\mathbf{I}\end{pmatrix}}_{\tilde{\mathbf{A}}}\begin{pmatrix}x(t)\\ z(t)\end{pmatrix} + \underbrace{\begin{pmatrix}\mathbf{B}\\ 0\end{pmatrix}}_{\tilde{\mathbf{B}}}u(t),\\ y(t) &= \underbrace{\begin{pmatrix}\mathbf{C} & 0\end{pmatrix}}_{\tilde{\mathbf{C}}}\begin{pmatrix}x(t)\\ z(t)\end{pmatrix}, \quad \begin{pmatrix}x(0)\\ z(0)\end{pmatrix} = \begin{pmatrix}x_0\\ 0\end{pmatrix}.\end{aligned} \tag{4.2}$$

In other words, replacing (1.1) by a reduced system of the form (1.3) can be interpreted as structure-preserving model reduction for a coupled problem. Applying the results from Proposition 3.4 to this case yields the following formulas for $\mathbf{P}$ and $\mathbf{Q}$:

$$\mathbf{P} = \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathbf{R}(\imath\omega)\mathbf{B}\mathbf{B}^T\mathbf{R}^*(\imath\omega)\,\mathrm{d}\omega, \quad \mathbf{Q} = \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathbf{R}^*(\imath\omega)\mathbf{C}^T\mathbf{C}\mathbf{R}(\imath\omega)\,\mathrm{d}\omega,$$

where $\mathbf{R}(\imath\omega) = (\imath\omega\mathbf{I} - \mathbf{A} - \frac{1}{\imath\omega+\gamma_1}\mathbf{K}_1)^{-1}$. On the other hand, for the coupled system, we have that

$$\tilde{\mathbf{P}} = \frac{1}{2\pi}\int_{-\infty}^{\infty} \tilde{\mathbf{R}}(\imath\omega)\tilde{\mathbf{B}}\tilde{\mathbf{B}}^T\tilde{\mathbf{R}}^*(\imath\omega)\,\mathrm{d}\omega, \quad \tilde{\mathbf{Q}} = \frac{1}{2\pi}\int_{-\infty}^{\infty} \tilde{\mathbf{R}}^*(\imath\omega)\tilde{\mathbf{C}}^T\tilde{\mathbf{C}}\tilde{\mathbf{R}}(\imath\omega)\,\mathrm{d}\omega.$$

In particular, considering the block structure

$$\tilde{\mathbf{R}}(\imath\omega) = \begin{pmatrix} \tilde{\mathbf{R}}_{11}(\imath\omega) & \tilde{\mathbf{R}}_{12}(\imath\omega) \\ \tilde{\mathbf{R}}_{21}(\imath\omega) & \tilde{\mathbf{R}}_{22}(\imath\omega) \end{pmatrix} = \left(\imath\omega\mathbf{I} - \begin{pmatrix} \mathbf{A} & \mathbf{I} \\ \mathbf{K}_1 & -\gamma_1\mathbf{I} \end{pmatrix}\right)^{-1}$$

we find that $\tilde{\mathbf{R}}_{11}(\imath\omega) = \mathbf{R}(\imath\omega)$. Hence, our balancing step can be interpreted as balancing the (1,1) block of the regular Gramians of the linear system (4.2).

4.2. Time fractional systems. As a second example, let us consider a measure μ of the form

$$\mu(\mathrm{d}s) = \frac{s^{\alpha-1}}{\Gamma(\alpha)}\mathbf{A}\,\mathrm{d}s,$$

where $0 < \alpha < 1$ and $\Gamma(\cdot)$ denotes the gamma function. In this case, we obtain a time fractional system of the form

$$\begin{aligned} \dot{x}(t) &= \frac{1}{\Gamma(\alpha)}\int_0^t s^{\alpha-1}\mathbf{A}x(t-s)\,\mathrm{d}s + \mathbf{B}u(t) \\ &= \frac{1}{\Gamma(\alpha)}\int_0^t (t-s)^{\alpha-1}\mathbf{A}x(s)\,\mathrm{d}s + \mathbf{B}u(t), \\ y(t) &= \mathbf{C}x(t), \quad x(0) = 0. \end{aligned} \tag{4.3}$$

Since for $0 < \alpha < 1$ it holds that $(\cdot)^{\alpha-1} \in L^1_{\mathrm{loc}}(0,\infty;\mathbb{R})$ we conclude that $\mu \in \mathcal{M}_{\mathrm{loc}}(0,\infty;\mathbb{R}^{n\times n})$ such that we may apply Theorem 2.4. Note that the integral term in (4.3) in fact denotes the Riemann–Liouville integral J^α of order α of the function $x(\cdot)$; see, e.g., [38]. With ∂ denoting time differentiation, instead of (4.3), we can write [13, 14]

$$\dot{x}(t) = \partial^{-\alpha}\mathbf{A}x(t) + \mathbf{B}u(t).$$

Provided the input signal is sufficiently smooth, this implies

$$\partial^{1+\alpha}x(t) = \mathbf{A}x(t) + \mathbf{B}(\partial^\alpha u(t)).$$

In other words, the specific choice of μ here results in time fractional systems of order $1 < \beta < 2$. In particular, if $\mathbf{A}$ is obtained by a spatial discretization of Δ, then $\beta = 1$ reflects the heat equation while $\beta = 2$ corresponds to the wave equation.

Let us address the computation of the Gramians $\mathbf{P}$ and $\mathbf{Q}$ as in Proposition 3.4. As we pointed out in Remark 3.5, the meaning of $\hat{\mu}(\imath\omega)$ is not clear a priori. On

the other hand, if the differential resolvent $\Phi(\cdot) \in L^2(0,\infty;\mathbb{R}^{n\times n})$ then its Fourier transform is well defined. In [13], systems of the form (4.3) have been considered in the more abstract setting

$$\dot{z}(t) = \frac{1}{\Gamma(\alpha)} \int_0^t (t-s)^{\alpha-1} \mathcal{A} z(s)\, \mathrm{d}s + f(t), \quad z(0) = z_0 \in \mathcal{Z},$$

where $\mathcal{A}\colon \mathcal{D}(\mathcal{A}) \subset \mathcal{Z} \to \mathcal{Z}$ is a linear, densely defined operator of sectorial type on a complex Banach space $\mathcal{Z}$ and $f\colon [0,T] \to \mathcal{Z}$. In particular, if $\mathcal{A}$ generates a semigroup with growth bound $\omega < 0$, it has been shown that the evolution operator $\mathcal{U}(t)$ satisfies $\|\mathcal{U}(t)\| \leq \frac{C}{1+|\omega|t^{1+\alpha}}$ for $t \geq 0$. For the example we consider in the next section, we thus find that $\Phi(\cdot) \in L^2(0,\infty;\mathbb{R}^{n\times n})$. Let us now come back to the finite-dimensional system (4.3). Following [38], for $z > 0$ applying the Laplace transform to (4.3), we obtain

$$\hat{x}(z) = (z\mathbf{I} - z^{-\alpha}\mathbf{A})^{-1}\mathbf{B}u(z)$$

and, therefore, the Laplace transform of the differential resolvent is $\hat{\Phi}(z) = (z\mathbf{I} - z^{-\alpha}\mathbf{A})^{-1}$. Due to our previous considerations and assumptions on the system (4.3), for $\mathbf{P}$ and $\mathbf{Q}$ we thus obtain

$$\mathbf{P} = \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathbf{R}(\imath\omega)\mathbf{B}\mathbf{B}^T\mathbf{R}^*(\imath\omega)\, \mathrm{d}\omega, \quad \mathbf{Q} = \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathbf{R}^*(\imath\omega)\mathbf{C}^T\mathbf{C}\mathbf{R}(\imath\omega)\, \mathrm{d}\omega,$$

where $\mathbf{R}(\imath\omega) = (\imath\omega\mathbf{I} - (\imath\omega)^{-\alpha}\mathbf{A})^{-1}$. We provide a corresponding example in the next section. We also point to [16], where a similar type of frequency-domain representation has been studied in the context of Maxwell equations.

4.3. Finite delay systems. We now focus on the relation between the results presented here and existing ones from the literature. We already pointed out similarities to the finite delay case. We thus expect to regain the Gramians from [25] when we choose

$$\mu(\mathrm{d}s) = \mathbf{A}\delta_0(\mathrm{d}s) + \mathbf{A}_1\delta_\tau(\mathrm{d}s).$$

Indeed, applying the Laplace transform to μ yields $\hat{\mu}(z) = \mathbf{A} + e^{-z\tau}\mathbf{A}_1$ and, consequently, $\mathbf{R}(z) = (z\mathbf{I} - \mathbf{A} - e^{-z\tau}\mathbf{A}_1)^{-1}$. Using a block inversion argument, the latter expression can be shown to coincide with the (1,1) block of the resolvent $(zI - \mathcal{A})^{-1}$ of the infinite-dimensional finite delay system. This relation has already been shown in [25, 37].

4.4. Second order systems. While we assumed μ to be a locally finite Borel measure, let us (only formally) set

$$\mu(\mathrm{d}s) = \mathbf{A}\delta_0(\mathrm{d}s) - \mathbf{M}\delta_0^{(2)}(\mathrm{d}s),$$

where $\delta_0^{(2)}$ denotes the second distributional derivative of the Dirac delta distribution. We then obtain the second order system

$$\begin{aligned} \mathbf{M}\ddot{x}(t) + \dot{x}(t) &= \mathbf{A}x(t) + \mathbf{B}u(t), \\ y(t) &= \mathbf{C}x(t), \quad x(0) = x_0,\ \dot{x}(0) = v_0. \end{aligned} \tag{4.4}$$

Again, we may rewrite the system in first order form as follows,

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix}x(t)\\ \dot{x}(t)\end{pmatrix} &= \underbrace{\begin{pmatrix}0 & \mathbf{I}\\ \mathbf{M}^{-1}\mathbf{A} & -\mathbf{M}^{-1}\end{pmatrix}}_{\tilde{\mathbf{A}}}\begin{pmatrix}x(t)\\ \dot{x}(t)\end{pmatrix} + \underbrace{\begin{pmatrix}0\\ \mathbf{M}^{-1}\mathbf{B}\end{pmatrix}}_{\tilde{\mathbf{B}}}u(t),\\ y(t) &= \underbrace{\begin{pmatrix}\mathbf{C} & 0\end{pmatrix}}_{\tilde{\mathbf{C}}}\begin{pmatrix}x(t)\\ \dot{x}(t)\end{pmatrix}, \quad \begin{pmatrix}x(0)\\ \dot{x}(0)\end{pmatrix} = \begin{pmatrix}x_0\\ v_0\end{pmatrix}.\end{aligned} \tag{4.5}$$

Let us compare the Gramians from Proposition 3.4 with those arising for second order balanced truncation (see, e.g., [11, 39]). Interpretation as a Volterra system (1.1) yields

$$\mathbf{P} = \frac{1}{2\pi}\int_{-\infty}^{\infty}\mathbf{R}(\imath\omega)\mathbf{B}\mathbf{B}^T\mathbf{R}^*(\imath\omega)\,\mathrm{d}\omega, \quad \mathbf{Q} = \frac{1}{2\pi}\int_{-\infty}^{\infty}\mathbf{R}^*(\imath\omega)\mathbf{C}^T\mathbf{C}\mathbf{R}(\imath\omega)\,\mathrm{d}\omega,$$

where $\mathbf{R}(\imath\omega) = (\imath\omega\mathbf{I} - \mathbf{A} + (\imath\omega)^2\mathbf{M})^{-1}$ while the classical Gramians of (4.5) are

$$\tilde{\mathbf{P}} = \frac{1}{2\pi}\int_{-\infty}^{\infty}\tilde{\mathbf{R}}(\imath\omega)\tilde{\mathbf{B}}\tilde{\mathbf{B}}^T\tilde{\mathbf{R}}^*(\imath\omega)\,\mathrm{d}\omega, \quad \tilde{\mathbf{Q}} = \frac{1}{2\pi}\int_{-\infty}^{\infty}\tilde{\mathbf{R}}^*(\imath\omega)\tilde{\mathbf{C}}^T\tilde{\mathbf{C}}\tilde{\mathbf{R}}(\imath\omega)\,\mathrm{d}\omega.$$

Using the position-velocity partitioning of the second order Gramians [39]

$$\tilde{\mathbf{P}} = \begin{pmatrix}\tilde{\mathbf{P}}_{\mathrm{p}} & \tilde{\mathbf{P}}_{12}\\ \tilde{\mathbf{P}}_{21} & \tilde{\mathbf{P}}_{\mathrm{v}}\end{pmatrix}, \quad \tilde{\mathbf{Q}} = \begin{pmatrix}\tilde{\mathbf{Q}}_{\mathrm{p}} & \tilde{\mathbf{Q}}_{12}\\ \tilde{\mathbf{Q}}_{21} & \tilde{\mathbf{Q}}_{\mathrm{v}}\end{pmatrix},$$

by simple algebraic manipulations, we can show that $\mathbf{P} = \tilde{\mathbf{P}}_{\mathrm{p}}$ and $\mathbf{Q} = \tilde{\mathbf{Q}}_{\mathrm{v}}$. Note that although the formal choice of μ is not explicitly covered by the results from section 3, it is well known that for *stable* second order systems all of the above integrals are well defined; see also [11, 39]. Hence, our approach results in the position-velocity balanced truncation method for second order systems.

5. Numerical examples. In this section, we study the performance of the suggested method by means of two different (spatially discretized) partial differential equations. We also describe a numerically efficient way of (approximately) computing the Gramians $\mathbf{P}$ and $\mathbf{Q}$ by means of a quadrature formula.

All simulations were generated on an Intel®Xeon(R) CPU E31270 @ 3.40 GHz x 8, 16 GB RAM, Ubuntu Linux 14.04, MATLAB Version 8.0.0.783 (R2012b) 64-bit (glnxa64).

5.1. Computing the Gramians. The most essential point in the construction of the reduced-order model (1.3) is the computation of the Gramians $\mathbf{P}$ and $\mathbf{Q}$. As we emphasized earlier, using the definition in (3.2) does not seem to be appealing since for large-scale systems the computation of the differential resolvent is prohibitively expensive. Instead, we recommend using the integral representations

$$\begin{aligned}\mathbf{P} &= \frac{1}{2\pi}\int_{-\infty}^{\infty}\mathbf{R}(\imath\omega)\mathbf{B}\mathbf{B}^T\mathbf{R}^*(\imath\omega)\,\mathrm{d}\omega = \frac{1}{\pi}\mathrm{Re}\left(\int_0^{\infty}\mathbf{R}(\imath\omega)\mathbf{B}\mathbf{B}^T\mathbf{R}^*(\imath\omega)\,\mathrm{d}\omega\right),\\ \mathbf{Q} &= \frac{1}{2\pi}\int_{-\infty}^{\infty}\mathbf{R}^*(\imath\omega)\mathbf{C}^T\mathbf{C}\mathbf{R}(\imath\omega)\,\mathrm{d}\omega = \frac{1}{\pi}\mathrm{Re}\left(\int_0^{\infty}\mathbf{R}^*(\imath\omega)\mathbf{C}^T\mathbf{C}\mathbf{R}(\imath\omega)\,\mathrm{d}\omega\right).\end{aligned}$$

For standard balanced truncation, approximating the Gramians via quadrature is sometimes referred to under the name *poor man's truncated balanced realization*; see [3, 36, 46]. Similarly, we are interested in a numerical approximation of the integrals

$$\int_0^\infty f(\omega)\,\mathrm{d}\omega := \int_0^\infty \mathbf{R}(\imath\omega)\mathbf{B}\mathbf{B}^T\mathbf{R}^*(\imath\omega)\,\mathrm{d}\omega,$$
$$\int_0^\infty \tilde{f}(\omega)\,\mathrm{d}\omega := \int_0^\infty \mathbf{R}(\imath\omega)^*\mathbf{C}^T\mathbf{C}\mathbf{R}(\imath\omega)\,\mathrm{d}\omega.$$

Let us focus on the first integral. Inspired by [7], one may, for example, use the Gauss–Kronrod quadrature formula. For this, we first transform the integration domain to $[-1,1]$ such that

$$\int_0^\infty f(\omega)\,\mathrm{d}\omega = \int_{-1}^1 f\left(\frac{\xi}{1-\xi}\right)\frac{1}{(1-\xi)^2}\,\mathrm{d}\xi = \int_{-1}^1 g(\xi)\,\mathrm{d}\xi.$$

For an adaptive computation, the latter integral is further divided (via bisection) into subintervals $[a,b] \subseteq [-1,1]$ and approximated via

$$\int_a^b g(\xi)\,\mathrm{d}\xi \approx \sum_{k=1}^q \nu_k g(\xi_k),$$

where the weights $\nu_k > 0$ and nodes ξ_k are obtained by the standard Gauss–Kronrod formula. For a more detailed description, we additionally refer to [7]. One benefit of using the Gauss–Kronrod quadrature is the fact that it includes a Gauss quadrature of lower order such that, without additional effort, an error estimate can be obtained by analyzing the difference of the two approximations. In our numerical experiments, we use a Gauss–Kronrod quadrature with 7 Gauss and 15 Gauss–Kronrod points and weights for each subinterval. The bisection then is only done when the error estimate exceeds the desired tolerance.

Let us further comment on a numerically efficient low-rank decomposition of the approximate Gramians. After the integrals have been evaluated for each subinterval, we obtain an approximation of the form

$$\mathbf{P} \approx \mathbf{P}_m = \frac{1}{\pi}\mathrm{Re}\left(\sum_{i=1}^d \nu_i g(\xi_i)\right).$$

Note that the number of inputs and outputs often satisfy $m, p \ll n$ such that evaluations of g and f are in a *low-rank* form

$$g(\xi_i) = \mathbf{Y}_i\mathbf{Y}_i^*, \quad \mathbf{Y}_i \in \mathbb{C}^{n\times m}.$$

Provided the number of evaluations d is such that $dm, dp \ll n$, we also obtain a low-rank decomposition for the approximate Gramian

$$\mathbf{P}_m = \mathbf{Z}\mathbf{Z}^*, \quad \mathbf{Z} \in \mathbb{C}^{n\times dm}.$$

Of course, it is not clear a priori if such a decomposition can be expected. However, for the standard case it is known that the singular values of the Gramians $\mathbf{P}$ and $\mathbf{Q}$ often decay reasonably fast to allow for efficient low-rank representations. While a theoretical analysis for the generalized case considered here is out of the scope

of this paper, we emphasize that this seems to be an important topic for future research. Particularly, since, in the numerical experiments (see below), such a decay can be observed. Moreover, alternative low-rank methods such as Krylov subspace or alternating directions implicit based approaches could be further studied; see, e.g. [6].

Remark 5.1. Note that the (approximate) computation of the Gramians via a quadrature can be interpreted as a frequency domain POD method. For the standard case, this has already been investigated in [46]. The crucial observation is that if the system is excited by an *impulsive input* $u(t) = \delta_0(t)$, the corresponding frequency response is given by $\hat{\mathbf{x}}(\omega) = (\imath\omega\mathbf{I} - \hat{\mu}(\imath\omega))^{-1}\mathbf{B}$. Hence, a POD projection onto the subspace associated with the dominant singular values of the snapshot matrix is implicitly related to the approximation of the Gramian $\mathbf{P}$. Similarly, snapshots generated by the dual system can be interpreted as approximating $\mathbf{Q}$.

5.2. The heat equation with fading memory. The first example arises within the context of heat conduction in materials with fading memory [31, 35]. For $\Omega = (0,1) \times (0,1)$, let us consider the following system

$$\begin{aligned} v_t(t,x) &= \Delta v(t,x) - \int_0^t \gamma(t-s)\Delta v(s,x)\mathrm{d}s + \chi_\omega u(t) \quad \text{in } (0,\infty)\times\Omega, \\ v(t,x) &= 0 \quad \text{in } (0,\infty)\times\Gamma, \qquad v(0,x) = 0 \quad \text{in } \Omega, \end{aligned} \tag{5.1}$$

where Γ denotes the boundary of Ω and for $\omega \subset \Omega$ the function χ_ω is defined by

$$\chi_\omega(x) := \begin{cases} 1 & \text{if } x \in \omega, \\ 0 & \text{otherwise} \end{cases}$$

such that for fixed t, the control is constant on ω. Further assume that a measurement of the form

$$v_{\mathrm{obs}}(\cdot) = \int_\Omega v(\cdot,x)\ \mathrm{d}x$$

is given. As pointed out in [31], for stable systems, we have that $1 - \int_0^\infty \gamma(y)\mathrm{d}s > 0$. For the control domain we set

$$\omega = [0.15, 0.25] \times [0.2, 0.3].$$

After spatial discretization by finite differences, we obtain the Volterra integro-differential system

$$\begin{aligned} \dot{z}(t) &= \mathbf{A}z(t) - \int_0^t \mathbf{K}(t-s)z(s)\mathrm{d}s + \mathbf{B}u(t), \\ y(t) &= \mathbf{C}z(t), \quad z(0) = 0, \end{aligned} \tag{5.2}$$

where $\mathbf{A} \in \mathbb{R}^{n\times n}, \mathbf{K}(t) = e^{-\gamma t}\mathbf{A} \in \mathbb{R}^{n\times n}, \mathbf{B} \in \mathbb{R}^n$, and $\mathbf{C} \in \mathbb{R}^{1\times n}$. Here, $\mathbf{A}$ denotes the discretization of the Dirichlet Laplacian and $n = 65536$ is the number of interior grid points. In Figures 1 and 2, numerical results for different values of γ and r are shown. Note that for larger values of γ, the influence of the integro term is reduced and the dynamics are similar to the standard heat equation; see Figure 2 (left), where the Bode plot of the transfer function $G(\imath\omega) = \mathbf{C}(\imath\omega\mathbf{I} - \mathbf{A} + \frac{1}{\imath\omega+\gamma}\mathbf{A})^{-1}\mathbf{B}$ is shown. In Figure 1 (right) and Figure 2 (right) we also show the $\mathcal{H}_2$-norm of the

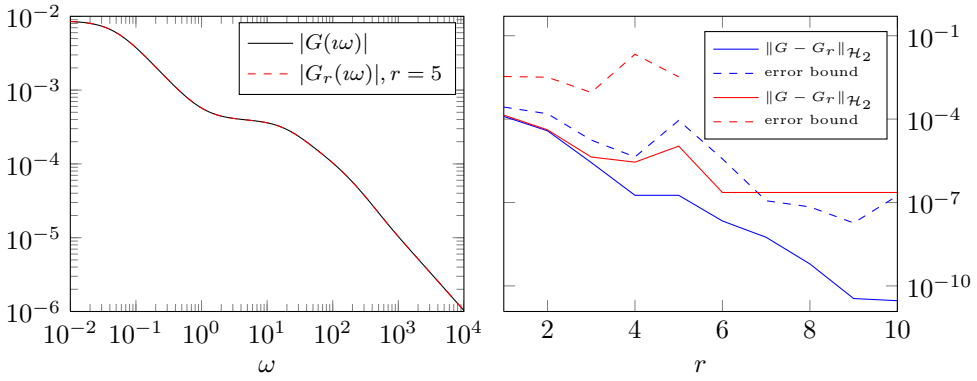

FIG. 1. *Heat equation with fading memory,* $\gamma = 1.05$, $n = 65536$.

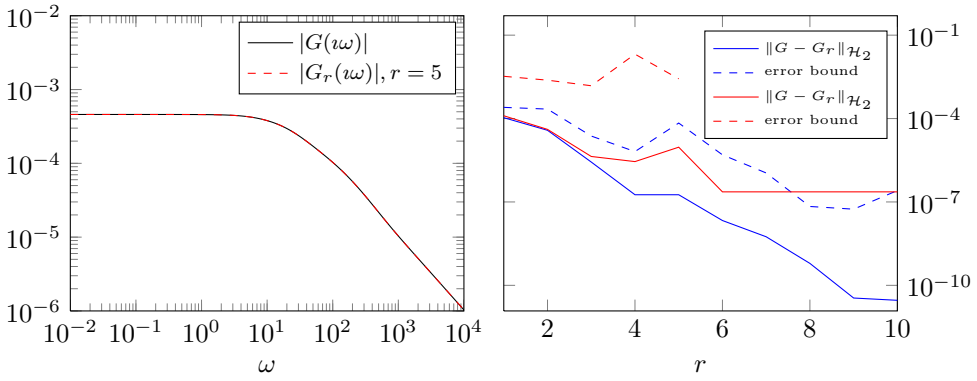

FIG. 2. *Heat equation with fading memory,* $\gamma = 10$, $n = 65536$.

error system for different reduced-system dimensions r. The Gramians are computed via a low-rank quadrature formula as described previously. The results are shown for approximations corresponding to two different accuracies. The estimated relative Frobenius norm error was 10^{-7} (blue) and 10^{-3} (red), respectively. Note that for the coarser approximation, the numerical rank of the Gramians was only 6. As a consequence, the corresponding $\mathcal{H}_2$-error bound will be zero for reduced-system dimensions $r \geq 6$ and larger reduced systems cannot be computed. On the other hand, for the finer approximation, the rank of the approximate Gramians was 19. As can be seen from the figures, in this case, the error bound as well as the reduced-order systems are more accurate.

5.3. A time fractional diffusion-wave equation. As a second example, we consider a time fractional diffusion-wave equation

$$\begin{aligned} \frac{\partial}{\partial t} v(t,x) &= \frac{1}{\Gamma(\alpha)} \int_0^t (t-s)^{\alpha-1} \Delta v(s,x) \mathrm{d}s + \chi_\omega u(t) \quad \text{in } (0,\infty) \times \Omega, \\ v(t,x) &= 0 \quad \text{in } (0,\infty) \times \Gamma, \qquad v(0,x) = 0 \quad \text{in } \Omega, \end{aligned} \tag{5.3}$$

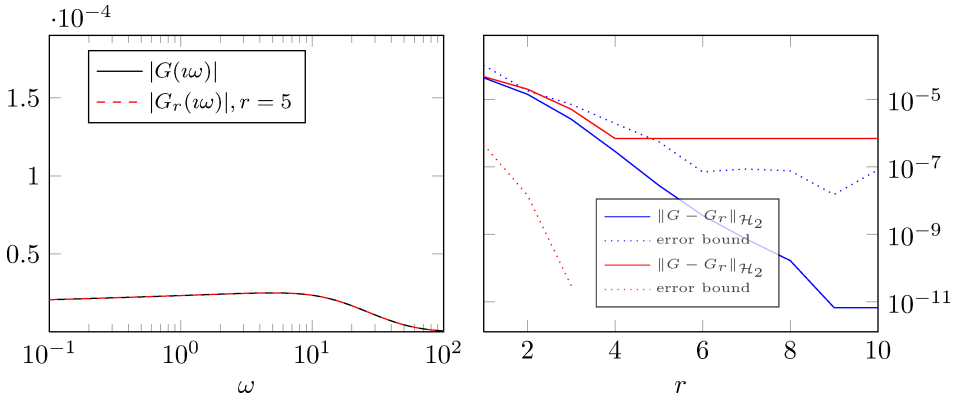

FIG. 3. *Time fractional diffusion-wave equation,* $\alpha = 0.05$, $n = 65536$.

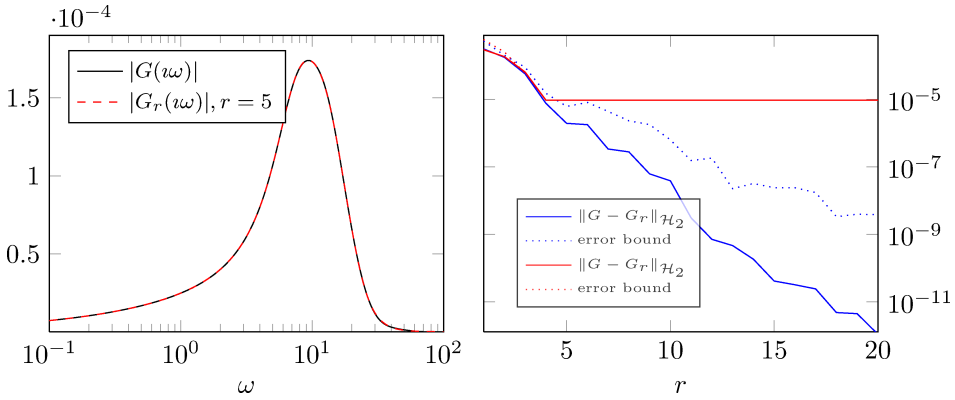

FIG. 4. *Time fractional diffusion-wave equation,* $\alpha = 0.5$, $n = 65536$.

where $0 < \alpha < 1$. The measurement now is assumed to be of the form

$$v_{\text{obs}}(\cdot) = \int_{\tilde{\omega}} v(\cdot, x)\ \mathrm{d}x.$$

For the control and observation domains we choose

$$\omega = [0.15, 0.35] \times [0.15, 0.35], \quad \tilde{\omega} = [0.65, 0.85] \times [0.65, 0.85].$$

A spatial discretization by finite differences then yields a system of the form (4.3). Again, the Gramians are computed for two different accuracies, indicated by blue (fine) and red (coarse).

In Figures 3 (left), 4 (left) and 5 (left) we compare the behavior of the system in terms of its transfer function $G(\imath\omega) = \mathbf{C}(\imath\omega\mathbf{I} - (\imath\omega)^{-\alpha}\mathbf{A})^{-1}\mathbf{B}$ for different fractional values α. As we expect, for $\alpha = 0.05$, the transfer function does not exhibit any peaks and behaves similarly to the one obtained for the standard heat equation. On the other hand, when α is increased, the transfer function becomes more and more irregular. In Figure 5 (left), we see the results for $\alpha = 0.95$. Since $\alpha = 1$ corresponds to the undamped wave equation, we expect the system poles to move closer to the imaginary axis. This is reflected in the increasing number of peaks of the transfer

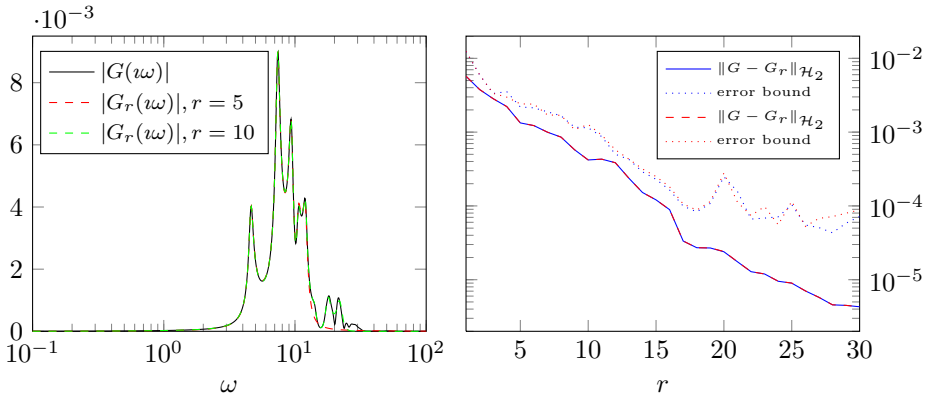

FIG. 5. *Time fractional diffusion-wave equation,* $\alpha = 0.95$, $n = 65536$.

function. A similar conclusion can be drawn from Figure 5 (right) where the decay of the error bound is significantly slower than for smaller values of α. Also, as in the first example, for smaller values of α, the coarse approximation yields approximate Gramians of very small rank such that a reliable error bound as well as reduced systems of larger dimension cannot be obtained; see Figures 3 (right) and 4 (right).

6. Conclusion. We proposed an extension of the method of balanced truncation to Volterra integro-differential equations. The approach relies on the simultaneous balancing of two system Gramians that share energy interpretations known from the standard LII case. The model reduction approach is structure-preserving and yields an error bound with respect to the $\mathcal{H}_2$-norm. We have discussed the relation to existing balancing based techniques that arise as we pointed out that one may use the method to reduce time fractional systems arising in the context of fractional diffusion-wave equations. Numerical examples show the applicability to large-scale systems arising from a spatial discretization of partial differential equations.

Acknowledgment. I would like to thank Karl Kunisch for several fruitful discussions on the subject and valuable comments for improvement of the presentation. Further, I would like to thank Peter Benner and Timo Reis for comments on a previous version of this manuscript.

REFERENCES

[1] A. ANTOULAS, *Approximation of Large-Scale Dynamical Systems*, Adv. Des., Control, SIAM, Philadelphia 2005.

[2] A. ANTOULAS, C. BEATTIE, AND S. GUGERCIN, *Interpolatory model reduction of large-scale dynamical systems*, in Efficient Modeling and Control of Large-Scale Systems, J. Mohammadpour and K. Grigoriadis, eds., Springer, New York, 2010, pp. 3–58.

[3] U. BAUR, P. BENNER, AND L. FENG, *Model order reduction for linear and nonlinear systems: A system-theoretic perspective*, Arch. Comput. Methods Eng., 21 (2014), pp. 331–358.

[4] C. BEATTIE AND S. GUGERCIN, *Interpolatory projection methods for structure-preserving model reduction*, Systems Control Lett., 58 (2009), pp. 225–232.

[5] P. BENNER, V. MEHRMANN, AND D. SORENSEN, *Dimension Reduction of Large-Scale Systems*, Lect. Notes Comput. Sci. Eng. 45, Springer, Berlin, 2005.

[6] P. BENNER AND J. SAAK, *Numerical solution of large and sparse continuous time algebraic matrix Riccati and Lyapunov equations: A state of the art survey*, GAMM-Mitt., 36 (2013), pp. 32–52.

[7] P. Benner and A. Schneider, *Balanced Truncation for Descriptor Systems with Many Terminals*, Technical report, Max Planck Institute, Magdeburg, Germany, 2013.
[8] J. A. Burns and T. L. Herdman, *Adjoint semigroup theory for a class of functional differential equations*, SIAM J. Math. Anal., 7 (1976), pp. 729–745.
[9] T. Burton, *Volterra Integral and Differential Equations*, Math. Sci. Eng. 202, Elsevier, Amsterdam, 2005.
[10] W. Castelan and E. Infante, *On a functional equation arising in the stability theory of difference-differential equations*, Quart. Appl. Math., 35 (1977), pp. 311–319.
[11] Y. Chahlaoui, D. Lemonnier, A. Vandendorpe, and P. V. Dooren, *Second-order balanced truncation*, Linear Algebra Appl., 415 (2006), pp. 373–384.
[12] A. Chatterjee, *An introduction to the proper orthogonal decomposition*, Current Sci., 78 (2000), pp. 808–817.
[13] E. Cuesta, *Asymptotic behaviour of the solutions of fractional integro-differential equations and some time discretizations*, Discrete Contin. Dyn. Syst., Suppl., 2007 (2007), pp. 277–285.
[14] E. Cuesta, C. Lubich, and C. Palencia, *Convolution quadrature time discretization of fractional diffusion-wave equations*, Math. Comp., 75 (2006), pp. 673–696.
[15] R. Curtain and H. Zwart, *An Introduction to Infinite-Dimensional Linear Systems Theory*, Springer, New York, 1995.
[16] L. Feng and P. Benner, *Model order reduction for systems with non-rational transfer function arising in computational electromagnetics*, in Scientific Computing in Electrical Engineering SCEE 2008, J. Roos and L. Costa, eds., Math. Ind., Springer, Berlin, 2010, pp. 515–522.
[17] K. Gallivan, A. Vandendorpe, and P. van Dooren, *Model reduction of MIMO systems via tangential interpolation*, SIAM J. Matrix Anal. Appl., 26 (2004), pp. 328–349.
[18] E. Grimme, *Krylov Projection Methods for Model Reduction*, PhD thesis, University of Illinois at Urbana-Champaign, 1997.
[19] G. Gripenberg, S.-O. Londen, and O. Staffans, *Volterra Integral and Functional Equations*, Encyclopedia Math. Appl., 34, Cambridge University Press, Cambridge, 1990.
[20] S. Grossman and R. Miller, *Nonlinear Volterra integrodifferential systems with L^1-kernels*, J. Differential Equations, 13 (1973), pp. 551–566.
[21] S. Gugercin and A. Antoulas, *A survey of model reduction by balanced truncation and some new results*, Internat. J. Control, 77 (2004), pp. 748–766.
[22] S. Gugercin, A. C. Antoulas, and C. Beattie, *$\mathcal{H}_2$ model reduction for large-scale linear dynamical systems*, SIAM J. Matrix Anal. Appl., 30 (2008), pp. 609–638.
[23] M. Heinkenschloss, T. Reis, and A. Antoulas, *Balanced truncation model reduction for systems with inhomogeneous initial conditions*, Automatica J. IFAC, 47 (2011), pp. 559–564.
[24] M. Hinze and S. Volkwein, *Proper orthogonal decomposition surrogate models for nonlinear dynamical systems: Error estimates and suboptimal control*, in Dimension Reduction of Large-Scale Systems, P. Benner, D. C. Sorensen, and V. Mehrmann, eds., Springer, Berlin, 2005, pp. 261–306.
[25] E. Jarlebring, T. Damm, and W. Michiels, *Model reduction of time-delay systems using position balancing and delay Lyapunov equations*, Math. Control Signals Systems, 25 (2013), pp. 147–166.
[26] V. Kharitonov, *Lyapunov matrices for a class of time delay systems*, Systems Control Lett., 55 (2006), pp. 610–617.
[27] V. Kharitonov and E. Plischke, *Lyapunov matrices for time-delay systems*, Systems Control Lett., 55 (2006), pp. 697–706.
[28] V. Kharitonov and A. Zhabko, *Lyapunov-Krasovskii approach to the robust stability analysis of time-delay systems*, Automatica J. IFAC, 39 (2003), pp. 15–20.
[29] K. Kunisch and S. Volkwein, *Galerkin proper orthogonal decomposition methods for parabolic problems*, Numer. Math., 90 (2001), pp. 117–148.
[30] K. Kunisch and L. Wang, *Time optimal controls of the linear FitzHugh–Nagumo equation with pointwise control constraints*, J. Math. Anal. Appl., 395 (2012), pp. 114–130.
[31] A. Lunardi, *On the linear heat equation with fading memory*, SIAM J. Math. Anal., 21 (1990), pp. 1213–1224.
[32] B. Moore, *Principal component analysis in linear systems: Controllability, observability and model reduction*, IEEE Trans. Automat. Control, 26 (1981), pp. 17–32.
[33] C. Mullis and R. Roberts, *Synthesis of minimum roundoff noise fixed point digital filters*, IEEE Trans. Circuits and Systems, 23 (1976), pp. 551–562.
[34] P. Ngoc, *Stability of Volterra-Stieltjes differential equations*, SIAM Journal on Control and Optimization, 49 (2011), pp. 205–226.

[35] J. Nunziato, *On heat conduction in materials with memory*, Quarterly of Applied Mathematics, (1971), pp. 187–304.
[36] J. Phillips and L. Silveira, *Poor man's TBR: a simple model reduction scheme*, IEEE Transaction on Computer-Aided Design of Integrated Circuits and Systems, 24 (2005), pp. 43–55.
[37] E. Plischke, *Transient Effects of Linear Dynamical Systems*, PhD thesis, Universität Bremen, 2005.
[38] I. Podlubny, *Fractional Differential Equations: An Introduction to Fractional Derivatives, Fractional Differential Equations, to Methods of Their Solution and Some of Their Applications*, vol. 198, Academic press, 1998.
[39] T. Reis and T. Stykel, *Balanced truncation model reduction of second-order systems*, Mathematical and Computer Modelling of Dynamical Systems, 14 (2008), pp. 391–406.
[40] C. Rowley, *Model reduction for fluids, using balanced proper orthogonal decomposition*, International Journal of Bifurcation and Chaos, 15 (2005), pp. 997–1013.
[41] L. Sirovich, *Turbulence and the dynamics of coherent structures*, Quarterly of Applied Mathematics, 45 (1987), pp. 561–571.
[42] D. Sorensen and A. Antoulas, *Dimension Reduction of Large-Scale Systems*, Springer Berlin Heidelberg, 2005, ch. On Model Reduction of Structured Systems, pp. 117–130.
[43] O. Staffans, *A direct Lyapunov approach to Volterra integrodifferential equations*, SIAM Journal on Mathematical Analysis, 19 (1998), pp. 879–901.
[44] C. De Villemagne and R. Skelton, *Model reductions using a projection formulation*, International Journal of Control, 46 (1987), pp. 2141–2169.
[45] V. Volterra, *Theory of Functionals and of Integral and Integro-Differential Equations*, Dover Publications, 2005.
[46] K. Willcox and J. Peraire, *Balanced model reduction via the proper orthogonal decomposition*, American Institute of Aeronautics and Astronautics Journal, 40 (2002), pp. 2323–2330.