Open access • Journal Article • DOI:10.1137/100797436

# Krylov-Based Model Order Reduction of Time-delay Systems — Source link

Wim Michiels, Elias Jarlebring, Karl Meerbergen

**Institutions:** Katholieke Universiteit Leuven

**Published on:** 01 Nov 2011 - SIAM Journal on Matrix Analysis and Applications (Society for Industrial and Applied Mathematics)

**Topics:** Discretization, Model order reduction, Eigenvalues and eigenvectors, Matrix (mathematics) and Reduction (complexity)

# KRYLOV BASED MODEL ORDER REDUCTION OF TIME-DELAY SYSTEMS

WIM MICHIELS , ELIAS JARLEBRING , AND KARL MEERBERGEN*

**Abstract.** We present a model order reduction method which allows the construction of a reduced, delay free model of a given dimension for linear time-delay systems, whose characteristic matrix is nonlinear due to the presence of exponential functions. The method builds on the equivalent representation of the time-delay system as an infinite-dimensional linear problem. It combines ideas from a finite-dimensional approximation via a spectral discretization on the one hand, and a Krylov-Padé model reduction approach on the other hand. The method exhibits a good spectral approximation of the original model, in the sense that the smallest characteristic roots are well approximated and the non-converged eigenvalues of the reduced model have a favorable location, and it preserves moments at zero and at infinity. The spectral approximation is due to an underlying Arnoldi process that relies on building an appropriate Krylov space for the linear infinite-dimensional problem. The preservation of moments is guaranteed, because the chosen finite-dimensional approximation preserves moments and, in addition, the space on which one projects is constructed in such a way that the preservation of moments carries over to the reduced model. The implementation of the method is dynamic, since the number of grid points in the spectral discretization does not need to be chosen beforehand and the accuracy of the reduced model can always be improved by doing more iterations. It relies on a reformulation of the problem involving a companion like system matrix and a highly structured input matrix, whose structure are fully exploited.

**Key words.** model reduction, Padé via Krylov, time-delay system

**1. Introduction.** The article concerns the development of reduced models for time-delay systems. Reduced models are useful in many situations and can, for instance, be used to reduce the computational cost for the solution in the time domain or the evaluation of the transfer function, or allow to use the reduced models in a control design. We consider a time-delay system of the form

$$\begin{cases} \dot{x}(t) &= A_0x(t) + \sum_{i=1}^m A_ix(t-\tau_i) + Bu(t), \\ y &= Cx(t) + Du(t), \end{cases} \tag{1.1}$$

where $x(t) \in \mathbb{C}^n$ is the state variable at time $t$, $u \in \mathbb{C}$ is the input, $y \in \mathbb{C}$ is the output and $\tau_i,\ i = 0, \ldots, m$, represent time-delays. We assume that

$$0 < \tau_1 < \cdots < \tau_m.$$

The transfer function of the system (1.1) is given by

$$\gamma(s) := C\left(sI - A_0 - \sum_{i=1}^m A_ie^{-s\tau_i}\right)^{-1} B + D. \tag{1.2}$$

The general problem we consider is to approximate the system (1.1) with a standard linear dynamical system without delay, in our context conveniently written as

$$\begin{cases} G\dot{z}(t) &= z(t) + Hu(t), \\ y(t) &= Fz(t) + Du(t), \end{cases} \tag{1.3}$$

where $z(t) \in \mathbb{C}^{k+1}$. As usual in a model reduction setting, we wish to find a reduced model of a given dimension, which is typically much smaller than the dimension of the

*K.U. Leuven, Department of Computer Science, 3001 Heverlee, Belgium, {Wim.Michiels,Elias.Jarlebring,Karl.Meerbergen}@cs.kuleuven.be

original model, i.e., $k \ll n$. More precisely, in this paper, we will present an efficient algorithm for computing $G$, $H$, $F$ and $D$ of small dimension such that the reduced system (1.3) approximates the original system (1.1) both in terms of approximation of the characteristic roots and in terms of derivatives of the transfer function at the origin and at infinity.

The problem we consider is hence the reduction of the infinite-dimensional system (1.1) to the system without delay (1.3). Results on these types of reduction techniques are rare in the literature, and, in fact, many problems related to such model reduction of time-delay systems are generally considered to be unsolved [21].

In the derivation of the proposed method we start by rewriting the system (1.1) as an equivalent infinite-dimensional linear system, as in [8]. The discretization of this system leads to a standard finite-dimensional linear problem, which is more suitable for model reduction purposes. The followed discretization approach is based on a spectral approximation, inspired by [5] where the corresponding eigenvalue problem was addressed. The accuracy depends on the choice of interpolation points in the interval $[-\tau_m, 0]$. Different choices of these points lead to different discretizations. We will choose the points in such a way that the accuracy of the eigenvalues is optimized and, at the same time, structure and sparsity can be introduced in the system matrices. Furthermore we prove that the transfer function of the discretized system matches several moments with the original transfer function (1.2). In the next step, where we project the (large) discretized system on a subspace, we guarantee that these moment matching properties are carried over to the reduced model by using a Padé via Krylov like model reduction method. In addition, by exploiting the structure of the problem during the construction of the Krylov space, as in [15], the process can be made dynamic in the sense that the number of discretization points in the spectral approximation does not need to be chosen be forehand, and the model reduction process can always be resumed if the accuracy of the reduced model is not sufficient.

For model reduction of linear systems based on moment matching, the Padé via Lanczos method and its variations are probably best known [9, 12, 11, 2]. These methods build a two-sided Krylov subspace with the system matrix and the input and output vectors as starting vectors. Krylov methods are also used for balanced truncation type of methods, see e.g. [13], and for parameterized model order reduction, see e.g. [1]. The advantage of two-sided methods is that both the input and the output are taken into account in the reduced models, which leads to matching twice as many moments for a given dimension of the reduced model compared to the case where only the input is taken into account. In many applications only one-sided methods are used. In particular, one-sided Krylov-Padé methods have become popular for the solution of large scale finite element models in structures and vibrations, which often lead to polynomial eigenvalue problems, see, e.g., [24, 3]. The proposed approach to derive a reduced model for (1.1) relies on a one-sided Krylov-Padé method, because this allows to fully exploit the structure of the problem. In addition, the proposed method leads to a good approximation of the smallest characteristic roots of the time-delay system. Since the rightmost characteristic roots are typically among the smallest ones [20], this makes the reduced models suitable for control design purposes. Two-sided methods use operations with the transpose of the matrices. Matrix operations with the transpose cannot exploit the special structure of the matrices, which implies that two sided methods do not preserve the structure and do not have the same dynamic properties as the one-sided method.

In the context sketched above, we recall that for linear systems of the form (1.3), the moments at zero are $-FH + D, -FGH, -2FG^2H, \ldots$ The moments at infinity, called Markov moments, are $D, FG^{-1}H, FG^{-2}H, \ldots$ In this paper, we will derive a reduced model from the Krylov sequence

$$\{G^{-1}H, H, GH, G^2H, \ldots G^{k-2}H\},$$

which matches the first two moments at infinity and $k-1$ moments at zero. The reason why moments are added at infinity is that the spectral discretization satisfies this property, and we want to keep it in the reduced model.

We note that Padé type model order reduction methods for nonlinear systems can also be used for (1.2), whose matrix is clearly nonlinear in $s$. The idea is to approximate the nonlinear system by a piecewise linear model, and then use a standard Krylov model reduction method on each piece separately [23]. This approach is not followed. Instead, we derive one reduced linear model, obtained — by approximation and projection — from a *linear* infinite-dimensional problem equivalent to (1.2).

We finally mention a series of results for a different type of approximation of time-delay systems. In [22, 16, 17] and derivative works, the authors propose a number of ways to approximate the time-delay system (1.1) by a finite-dimensional system and at the same time preserve important properties of the system. The approach can mostly be interpreted as a rational approximation of the exponential or the exponential times a rational function. Hence, if applied to the system (1.1) the corresponding reduced system does not contain a delay, but is however normally larger than the original system. In our approach we wish to find a reduced model (1.3) of a prescribed dimension, $k$, where typically $k \ll n$. There is also an interpolatory model reduction method applicable to time-delay systems[4].

The structure of the paper is as follows. In Section 2, we formulate the equivalent infinite-dimensional problem, outline the spectral discretization and discuss the properties of the discretized system, with the emphasis on moment matching. In Section 3 we present the dynamic Arnoldi method for the infinite-dimensional linear systems, and show how moment matching properties can be imposed on the reduced model. In Section 4, we illustrate the method and its properties by means of a numerical example. Finally, we formulate the main contributions of the paper in Section 5.

**2. Finite-dimensional approximation.** An approach to analyze the time-delay system (1.1) is to rewrite it in a linear infinite-dimensional form. The corresponding operators can be discretized, yielding an approximation of (1.1), involving large matrices and no delays. We will use here a spectral discretization, which is known to have appealing approximation properties. The discretization, is briefly summarized in the next paragraph. In §2.2 we discuss some properties of the discretized problem. As the main result of the section we prove that the discretized problem fulfills a moment matching property, which will play an important role in the derivation of the model reduction technique.

**2.1. A spectral discretization.** Consider the space $X := \mathbb{C}^n \times \mathcal{L}_2([-\tau_m, 0], \mathbb{C}^n)$, equipped with the inner product

$$< (y_0, y_1), (z_0, z_1) >_X = < y_0, z_0 >_{\mathbb{C}^n} + < y_1, z_1 >_{\mathcal{L}_2} .$$

We can now rewrite (1.1) as

$$\left\{ \begin{array}{rcl} \dot{z}(t) & = & \mathcal{A}z(t) + \mathcal{B}u(t), \\ y(t) & = & \mathcal{C}z(t) + Du(t), \end{array} \right. \tag{2.1}$$

where $\mathcal{A}:\ X \to X$ is a derivative operator defined by

$$\begin{aligned} \mathcal{D}(\mathcal{A}) = \{z = (z_0, z_1) \in X : z_1 \in \mathcal{C}([-\tau_m,\ 0],\ \mathbb{C}^n), \\ z_1' \in \mathcal{C}([-\tau_m, 0], \mathbb{C}^n),\ z_0 = z_1(0)\}, \\ \mathcal{A}z = \begin{pmatrix} A_0 z_0 + \sum_{i=1}^m A_i z_1(-\tau_i) \\ z_1' \end{pmatrix},\ z \in \mathcal{D}(\mathcal{A}) \end{aligned} \tag{2.2}$$

and the operator $\mathcal{B}:\ \mathbb{C} \to X$ and $\mathcal{C}:\ X \to \mathbb{C}$ are given by

$$\mathcal{B}u = \begin{pmatrix} u \\ 0 \end{pmatrix}, u \in \mathbb{C},\ \ \mathcal{C}z = Cz_0,\ z \in X.$$

This is a standard procedure to rewrite the time-delay system (1.1) as an infinite-dimensional system, see [8].

The relation between the solutions of (1.1) and (2.1) is $z_0(t) \equiv x(t)$, $z_1(t) \equiv x(t+\theta), \theta \in [-\tau_m, 0]$. We refer to [20, Chapter 1] for a detailed description of the spectral properties of the operator $\mathcal{A}$. Important here is that the operator only has a point spectrum and the eigenvalues of the operator $\mathcal{A}$ are equal to the characteristic roots of (1.1), i.e. the zeros of the characteristic equation,

$$\det\left(sI - A_0 - \sum_{i=1}^m A_i e^{-s\tau_i}\right) = 0. \tag{2.3}$$

We outline how the system (2.1) can be discretized using a *spectral method* (see, e.g. [25, 5]). Given a positive integer $N$, we consider a mesh $\Omega_N$ of $N+1$ distinct points in the interval $[-\tau_m,\ 0]$:

$$\Omega_N = \{\theta_{N,i},\ i = 1, \ldots, N+1\}, \tag{2.4}$$

where

$$-\tau_m \le \theta_{N,1} < \ldots < \theta_{N,N} < \theta_{N,N+1} = 0.$$

This allows to replace the continuous space $X$ with the space $X_N$ of discrete functions defined over the mesh $\Omega_N$, i.e. any function $\phi \in X$ is discretized into a block vector $x = [x_1^T \cdots\ x_{N+1}^T]^T \in X_N$ with components

$$x_i = \phi(\theta_{N,i}) \in \mathbb{C}^n,\ \ i = 1, \ldots, N+1.$$

We let $\mathcal{P}_N x,\ x \in X_N$, be the unique $\mathbb{C}^n$ valued interpolating polynomial of degree smaller than or equal to $N$, satisfying

$$\mathcal{P}_N x(\theta_{N,i}) = x_i,\ \ i = 1, \ldots, N+1.$$

In this way we can approximate the operator $\mathcal{A}$ over $X$ with the matrix $A_N:\ X_N \to X_N$, defined as

$$\begin{cases} (\mathcal{A}_N\ x)_i = & (\mathcal{P}_N x)'(\theta_{N,i}), & i = 1, \ldots, N, \\ (\mathcal{A}_N\ x)_{N+1} = & A_0 \mathcal{P}_N x(0) + \sum_{i=1}^m A_i \mathcal{P}_N x(-\tau_i). \end{cases} \tag{2.5}$$

Using the Lagrange representation of $\mathcal{P}_N x$,

$$\mathcal{P}_N x = \sum_{k=1}^{N+1} l_{N,k}\ x_k,$$

where the Lagrange polynomials $l_{N,k}$ are real valued polynomials of degree $N$ satisfying

$$l_{N,k}(\theta_{N,i}) = \begin{cases} 1 & i = k, \\ 0 & i \neq k, \end{cases}$$

we get an explicit form for the matrix $A_N$,

$$A_N = \begin{bmatrix} d_{1,1} & \dots & d_{1,N+1} \\ \vdots & & \vdots \\ d_{N,1} & \dots & d_{N,N+1} \\ a_1 & \dots & a_{N+1} \end{bmatrix} \in \mathbb{R}^{(N+1)n \times (N+1)n}, \tag{2.6}$$

where

$$\begin{cases} d_{i,k} &= l'_{N,k}(\theta_{N,i}) I_n, & i \in \{1, \dots, N\},\ k \in \{1, \dots, N+1\}, \\ a_k &= A_0 l_{N,k}(0) + \sum_{i=1}^m A_i l_{N,k}(-\tau_i), & k \in \{1, \dots, N+1\}. \end{cases}$$

In the same way we can approximate $\mathcal{B}$ and $\mathcal{C}$ by

$$B_N = [0\ \cdots\ 0\ 1]^T \otimes B, \quad C_N = [0\ \cdots\ 0\ 1] \otimes C$$

and we arrive at the finite-dimensional approximation of (1.1),

$$\begin{cases} \dot{z}(t) = A_N z(t) + B_N u(t),\ z(t) \in \mathbb{R}^{(N+1)n \times 1}, \\ y(t) = C_N z(t) + D u(t). \end{cases} \tag{2.7}$$

Accordingly, we can approximate (1.2) by the transfer function of (2.7), given by

$$\gamma_N(s) := C_N \left(sI - A_N\right)^{-1} B_N + D. \tag{2.8}$$

**2.2. Properties.** The discretization of the operator formulation of the time-delay system in the previous section resulted in the construction of a large standard dynamical system (2.7). It is natural to expect that the discretized system approximates the time-delay system. We now see that apart from the expected approximation properties of the spectrum the approximation also automatically fulfills a moment matching property. Several derivatives at the origin and the first derivative at infinity of the transfer function original (1.2) and the corresponding approximation (2.8) coincide.

THEOREM 2.1. *The transfer functions (1.2) and (2.8) satisfy,*

$$\left.\frac{d^i \gamma_N(s)}{ds^i}\right|_{s=0} = \left.\frac{d^i \gamma(s)}{ds^i}\right|_{s=0}, \quad i = 0, \dots, N, \tag{2.9}$$

*and*

$$\left.\frac{d^i \gamma_N(s^{-1})}{ds^i}\right|_{s=0} = \left.\frac{d^i \gamma(s^{-1})}{ds^i}\right|_{s=0}, \quad i = 0, 1, \tag{2.10}$$

*that is, the moments of $\gamma(s)$ and $\gamma_N(s)$ at zero match up to the $N$th moment, and the moments at infinity match up to the first moment.*

*Proof.* We first prove (2.9). In [14] it is shown that

$$\gamma_N(s) = C \left( sI - A_0 - \sum_{i=1}^m A_i p_N(-\tau_i;\ s) \right)^{-1} B + D, \tag{2.11}$$

where the function

$$t \in \mathbb{R} \mapsto p_N(t;\ s)$$

is the (unique) polynomial of degree $N$ satisfying

$$\begin{cases} p_N(0;\ s) = 1, \\ p'_N(\theta_{N,i};\ s) = s p_N(\theta_{N,i};\ s), \quad i \in \{-N, \dots, -1\}. \end{cases} \tag{2.12}$$

In other words, the effect of the approximating (2.1) by (2.7) can be interpreted in the frequency domain as the effect of approximating the exponential function $e^{-s\tau_i}$ in (1.2) by the function $p_N(-\tau_i;\ s)$ for $i = 0, \dots, N$.

Let us express $p_N$ in the monomial basis,

$$p_N = \sum_{i=0}^{N} c_i t^i.$$

The conditions (2.12) lead to

$$(sM - N) \begin{bmatrix} c_0 \\ \vdots \\ c_N \end{bmatrix} = F,$$

where

$$M = \begin{bmatrix} 1 & \theta_1 & \theta_1^2 & \cdots & \theta_1^N \\ \vdots & & & & \vdots \\ 1 & \theta_N & \theta_N^2 & \cdots & \theta_N^N \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix}, \quad N = \begin{bmatrix} 0 & 1 & 2\theta_1 & 3\theta_1^2 & \cdots & N\theta_1^{N-1} \\ \vdots & & & & \vdots & \\ 0 & 1 & 2\theta_N & 3\theta_N^2 & \cdots & N\theta_N^{N-1} \\ 1 & 0 & 0 & 0 & \cdots & 0 \end{bmatrix},$$

and $F = [0\ \cdots\ 0\ -1]^T$. In this way we get

$$p_N(t;\ s) = [1\ t\ t^2\ \cdots\ t^N](sM - N)^{-1} F$$

and

$$\left.\frac{\partial^i p_N(t;\ s)}{\partial s^i}\right|_{s=0} = -i!\ [1\ t\ t^2\ \cdots\ t^N]\ N^{-1}(MN^{-1})^i F, \quad i = 0, \dots, N.$$

By direct inspection we have

$$-N^{-1}F = [1\ 0\ \cdots\ 0]^T,$$

hence,

$$\begin{aligned} -N^{-1}(MN^{-1})F &= N^{-1}M\ [1\ 0\ \cdots\ 0]^T \\ &= N^{-1}\ [1\ \cdots\ 1\ 0]^T = [0\ 1\ 0\ \cdots\ 0]^T. \end{aligned}$$

By applying the same steps we get

$$-i!\ N^{-1}((MN)^{-1})^i F = e_i, \quad i = 0, \dots, N,$$

where $e_i$ is the $i$th unity vector in $\mathbb{C}^{N+1}$. We conclude that

$$\left.\frac{\partial^i p_N(t;\ s)}{\partial s^i}\right|_{s=0} = t^i = \left.\frac{\partial^i \left(e^{st}\right)}{\partial s^i}\right|_{s=0}, \quad i = 0, \ldots, N.$$

This result on its turn implies

$$\left.\frac{\partial^i p_N(-\tau_k;\ s)}{\partial s^i}\right|_{s=0} = \left.\frac{\partial^i \left(e^{-s\tau_k}\right)}{\partial s^i}\right|_{s=0}, \quad i = 0, \ldots, N,\ k = 1, \ldots, m. \tag{2.13}$$

Taking into account the equality (2.11), an application of the chain rule leads to the assertion (2.9)

Assertion (2.10) follows by a direct computation, as both the left and right hand size of (2.10) are equal to $CB$ for $i = 1$ and $D$ for $i = 0$. □

It is important to note that the properties described by Theorem 2.1 are *independent* of the choice of the grid points. Hence, other desired properties can be imposed by an optimal choice of the distribution of the grid points.

In what follows we choose the nonzero grid points as scaled and shifted zeros of $U_N$, the Chebyshev polynomial of the second kind and order $N$, i.e. the grid points are specified as

$$\theta_{N,i} = \frac{\tau_m}{2}(\alpha_{N,i} - 1), \quad \alpha_{N,i} = -\cos\frac{\pi i}{N+1},\ i = 1, \ldots, N+1. \tag{2.14}$$

With the choice of the Chebyshev grid (2.14) the convergence of the individual eigenvalues of $A_N$ to corresponding characteristic roots is fast. More specifically, in [5] it is proven that spectral accuracy (approximation error $O(N^{-N})$) is obtained[1].

An additional property of using a Chebyshev grid, observed in extensive numerical simulations, is that the eigenvalues of $A_N$, which have not yet converged to corresponding characteristic roots, are located to the left of the eigenvalues that have already converged (see, e.g., the plots in [5]). This property, which is important in the context of stability assessment, is illustrated with the following example.

EXAMPLE 2.2. *In Figure 3.1 we show the rightmost characteristic roots of the scalar time-delay system*

$$\dot{x}(t) = -x(t) - x(t-1), \tag{2.15}$$

*as well as the corresponding eigenvalues of the matrix (2.6), obtained by a discretization using the grid points (2.14), for $N = 9$ and $N = 19$. The eigenvalues of $A_N$, which have not yet converged to characteristic roots are located to the left of the converged eigenvalues.*

Other arguments for choosing the grid points (2.14) are given in the following section.

[1] In [5] a grid of (scaled and shifted) Chebyshev extremal points is used, the latter given by $\alpha_{N,i} = -\cos\frac{\pi i}{N},\ i = 0, \ldots, N$. The slight difference with (2.14) does, however, not affect the property of spectral converges, because the asymptotic *distribution* of the grid points, which determines the convergence properties [25], is the same for both grids.

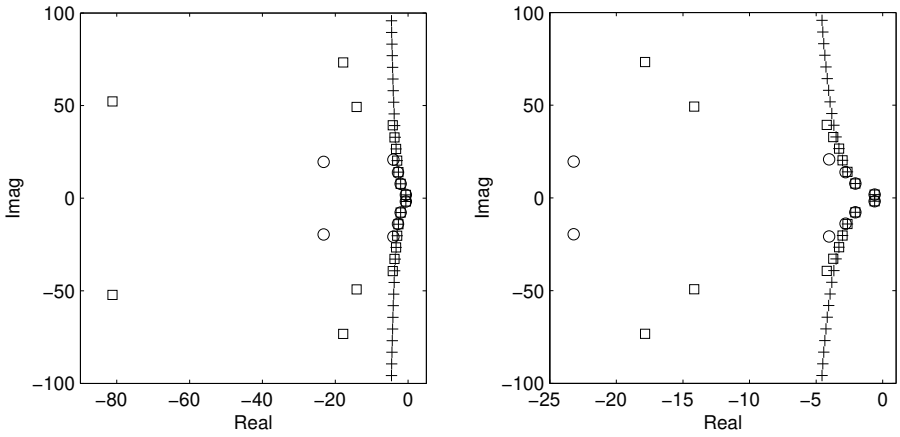



FIG. 2.1. *The rightmost characteristic roots of (2.15) are indicated with ('+'). The eigenvalues of $A_N$ are shown for $N = 9$ (indicated with 'o' -approximation by a 10-by-10 matrix) and $N = 19$ (indicated with '□' - approximation by a 20-by-20 matrix). Note that the difference between (a) and (b) is the range of the x-axis.*

**3. Constructing a reduced-order model.** We now know that the discretized system (2.7) has the nice approximation property that many moments are matched. It is however not a solution to our main problem to construct a small reduced model since the state space dimension of (2.7) is $n(N + 1)$, i.e., much larger than the state space dimension of the original time-delay system (1.1).

However, unlike the original time-delay system (1.1), the discretized system (2.7) is a standard dynamical system. We could hence conceptually reduce the dimension of the discretized system by applying a standard Krylov based model reduction technique on (2.7). This would involve explicitly constructing the large matrices in (2.7). We will now see that this can be avoided and an efficient implementation becomes possible, where many properties of (2.7) are exploited. Moreover, the construction is dynamic in the sense that the value of $N$ in (2.8) does not need to be fixed beforehand.

More specifically, in §3.1 we derive an equivalent representation of (2.7) and (2.8), where the matrices have a sparse structure. In §3.2 we dynamically construct a Krylov space. In §3.3 we project the system matrices on this subspace and outline how moment matching properties can be guaranteed. Finally, in §3.4 we discuss the various aspects of the resulting model reduction procedure.

The technical derivation of the results makes use of the representation of polynomials related to the spectral discretization in appropriately defined Chebyshev bases. In what follows we denote by $T_i$ the Chebyshev polynomial of the first kind and order $i$, and $U_i$ is the Chebyshev polynomial of the second kind and order $i$, with $i \geq 0$.

**3.1. A sparse reformulation of the problem.** In the derivation we will use a slightly different formulation of the discretization. Note that the eigenvalue problem

$$(sI - A_N)x = 0, \quad x \in \mathbb{C}^{(N+1)n}, \ x \neq 0, \tag{3.1}$$

where $A_N$ is given by (2.5), can directly be obtained by requiring that there exists a polynomial of degree $N$,

$$(\mathcal{P}_N x)(t) = \sum_{k=0}^{N} l_{N,k}(t)\ x_k,$$

which satisfies the conditions

$$\begin{cases} s\mathcal{P}_N x(\theta_{N,i}) = (\mathcal{P}_N x)'(\theta_{N,i}), & i \in \{1,\dots,N\}, \\ s\mathcal{P}_N x(0) = A_0 \mathcal{P}_N x(0) + \sum_{i=1}^{m} A_i \mathcal{P}_N x(-\tau_i). \end{cases} \tag{3.2}$$

Then the vector $x$ in (3.1) is obtained as $x = [x_0^T\ \cdots\ x_N^T]^T$. Hence, an eigenvalue problem equivalent to (3.1) can be obtained by expressing $\mathcal{P}_N x$ in another basis and imposing the same conditions. We now represent $\mathcal{P}_N x$ in a basis of Chebyshev polynomials:

$$(\mathcal{P}_N x)(t) = \sum_{i=0}^{N} c_i T_i\left(2\frac{t}{\tau_m} + 1\right), \tag{3.3}$$

where $T_i$ is the Chebyshev polynomial of the first kind and order $i$, and $c_i \in \mathbb{C}^{N\times 1}$ for $i = 0,\dots,N$. By requiring that this polynomial satisfies the conditions (3.2) we obtain an equivalent sparse eigenvalue problem for (3.1), as expressed in the following lemma (Theorem 2.1 from [15]).

LEMMA 3.1. *If the grid points in the spectral discretization of (2.1) are chosen as (2.14), then the eigenvalue problem (3.1) is equivalent with*

$$(s\Pi_N - \Sigma_N)\, c = 0,\ s \in \mathbb{C},\ c \in \mathbb{C}^{(N+1)n},\ c \neq 0, \tag{3.4}$$

*where*

$$\Pi_N = \frac{\tau_m}{4} \begin{bmatrix} \frac{4}{\tau_m} & \frac{4}{\tau_m} & \frac{4}{\tau_m} & \cdots & & \cdots & \frac{4}{\tau_m} \\ 2 & 0 & -1 & & & & \\ & \frac{1}{2} & 0 & -\frac{1}{2} & & & \\ & & \frac{1}{3} & 0 & \ddots & & \\ & & & \frac{1}{4} & \ddots & -\frac{1}{N-2} & \\ & & & & \ddots & 0 & -\frac{1}{N-1} \\ & & & & & \frac{1}{N} & 0 \end{bmatrix} \otimes I \tag{3.5}$$

*and*

$$\Sigma_N = \begin{bmatrix} R_0 & R_1 & \cdots & R_N \\ & I_n & & \\ & & \ddots & \\ & & & I_n \end{bmatrix}, \tag{3.6}$$

*with*

$$R_i = A_0 T_i(1) + \sum_{k=1}^{m} A_k T_i\left(-2\frac{\tau_k}{\tau_m} + 1\right),\ i = 0,\dots,N.$$

A comparison between (3.1) and (3.4), taking into account the interpretation of the vectors $c$ and $d$ as coefficients in polynomial bases, learns that

$$A_N = (S_N \otimes I)(\Pi_N^{-1}\Sigma_N)(S_N^{-1} \otimes I), \tag{3.7}$$

where the matrix $S_N \in \mathbb{R}^{(N+1)\times(N+1)}$ maps coefficients of a polynomial of degree $N$ in the Chebyshev basis

$$\left\{ T_i\left(2\frac{t}{\tau_m}+1\right) : \ i = 0, \ldots, N \right\} \tag{3.8}$$

onto the corresponding coefficients in the Lagrange basis,

$$\{l_{N,i}(t) : \ i = 1, \ldots, N+1\},$$

defined on the grid (2.14). The relation (3.7) leads to an alternative formulation of the transfer function $\gamma_N(s)$, as expressed in the following theorem.

THEOREM 3.2. *If the grid points in the spectral discretization of (2.1) are chosen as (2.14), then we can express*

$$\gamma_N(s) = F_N \left(sG_N - I\right)^{-1} H_N + D, \tag{3.9}$$

*where*

$$G_N = \Sigma_N^{-1}\Pi_N, \tag{3.10}$$

$$H_N = \begin{bmatrix} R_0^{-1}\left(I - \frac{\tau_m}{2}R_1\right)R_0^{-1}B \\ \frac{\tau_m}{2}R_0^{-1}B \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{3.11}$$

*and*

$$F_N = [CR_0 \ CR_1 \ \cdots \ CR_N], \tag{3.12}$$

*with $\Sigma_N$ and $\Pi_N$ defined by (3.6) and (3.5).*

*Proof.* Substituting (3.7) into (2.8) we immediately obtain (3.10). Moreover we can bring (2.8) in the form (3.9), where

$$H_N = G_N^2(S_N^{-1} \otimes I)B_N, \tag{3.13}$$

$$F_N = C_N(S_N \otimes I)G_N^{-1}. \tag{3.14}$$

In order to prove (3.11) we first find an explicit expression for

$$c := S_N^{-1}[0 \ \cdots \ 0 \ 1]^T. \tag{3.15}$$

From the definition of $S_N$ it follows that the elements of $c = [c_0 \ \cdots \ c_N]^T$ are the coefficients of the Lagrange polynomial $l_{N,N+1}$, expressed in the Chebyshev basis (3.8). Therefore, they satisfy the conditions

$$\begin{bmatrix} 1 & T_1(\alpha_{N,1}) & \cdots & T_N(\alpha_{N,1}) \\ \vdots & & & \vdots \\ 1 & T_1(\alpha_{N,N}) & \cdots & T_N(\alpha_{N,N}) \\ 1 & 1 & \cdots & 1 \end{bmatrix} \begin{bmatrix} c_0 \\ \vdots \\ c_{N-1} \\ c_N \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}. \tag{3.16}$$

When substituting

$$c = \begin{cases} \frac{2}{N+1}\ [0\ 1\ 0\ 1\ \cdots\ 0\ 1]^T, & N \text{ odd}, \\ \frac{2}{N+1}\ [\frac{1}{2}\ 0\ 1\ 0\ 1\ \cdots\ 0\ 1]^T, & N \text{ even} \end{cases} \tag{3.17}$$

and taking into account the property

$$T_0(t) = U_0(t),\ T_1(t) = \frac{1}{2}U_1(t),\ \ T_i(t) = \frac{1}{2}(U_i(t) - U_{i-2}(t)),\ i \geq 2,$$

the left hand side of (3.16) becomes

$$\frac{1}{N+1}[U_N(\alpha_{N,1})\ \cdots\ U_N(\alpha_{N,N})\ N+1]^T.$$

By the choice of the grid points (2.14) we see that (3.17) solves (3.16). This leads us to

$$(S_N^{-1} \otimes I)B_N = c \otimes B = \begin{cases} \frac{2}{N+1}\ [0\ 1\ 0\ 1\ \cdots\ 0\ 1]^T \otimes B, & N \text{ odd}, \\ \frac{2}{N+1}\ [\frac{1}{2}\ 0\ 1\ 0\ 1\ \cdots\ 0\ 1]^T \otimes B, & N \text{ even}. \end{cases} \tag{3.18}$$

When substituting this in (3.13), we get

$$H_N = G_N(G_N(c \otimes B)) = G_N[(R_0^{-1}B)^T\ 0\ \cdots\ 0]^T.$$

A direct computation yields (3.11).

In order to prove (3.12) we let $q = [q_0\ q_1\ \cdots\ q_N]^T \in \mathbb{C}^{(N+1)\otimes 1}$. Then we get

$$[0\ \cdots\ 0\ 1]\ S_N\ q = \sum_{i=0}^{N} q_i T_i\left(2\frac{\theta_{N,N+1}}{\tau_m} + 1\right) = [1\ 1\ \cdots\ 1]q.$$

In this way (3.14) becomes

$$F_N = [C\ \cdots\ C]G_N^{-1} = [C\ \cdots\ C]\Pi_N^{-1}\Sigma_N = [C\ 0\ \cdots\ 0]\Sigma_N.$$

The expression (3.12) follows. □

The following property of the matrices in (3.9) plays an important role in the next paragraphs.

PROPOSITION 3.3. *Assume that $N_1, N_2 \in \mathbb{N}$ with $N_1 < N_2$. Then the matrices $\Sigma_{N_1}, \Pi_{N_1}, F_{N_1}, H_{N_1}$ in Theorem 3.2 are submatrices of $\Sigma_{N_2}, \Pi_{N_2}, F_{N_2}, H_{N_2}$.*

Proposition 3.3 allows an adaptive construction of the approximation. An increase of the number of grid points, $N$, can be dealt with by extending the corresponding matrices..

**3.2. Dynamic construction of a Krylov space.** The model reduction technique presented in the paper is based on projecting the large and sparse matrices $F_N$, $G_N$ and $H_N$, defined in Theorem 3.2, on an appropriately defined subspace. Instrumental to this we use the dynamic construction of a Krylov space of $G_N$, presented in [15]. This construction on its turn is inspired by methods for polynomial eigenvalue problems that exploit structure to reduce the storage cost of the Krylov vectors [3, 10, 18]. In what follows we summarize this construction (in a slightly adapted form). In §3.3, we derive reduced models based on a projection on the Krylov subspace.

We fix integers $k$ and $N$, satisfying $1 \leq k \leq N$ and consider the Krylov space

$$\mathcal{K}_k(G_N, b) := \text{span}\{b, G_N b, \ldots, G_N^{k-1} b\}. \tag{3.19}$$

Due to the special structure of $G_N$ the matrix vector product satisfies the following property [15, Theorem 3.1].

THEOREM 3.4. *Assume that* $(\sum_{i=0}^{m} A_i)$ *is non-singular. Let* $i, N \in \mathbb{N}$ *with* $i \leq N$ *and let* $Y \in \mathbb{C}^{n \times i}$. *Then*

$$G_N \text{vec}(Y, 0, 0, \ldots, 0) = \text{vec}(\hat{x}, Z, 0, \ldots, 0),$$

*where* $Z \in \mathbb{C}^{n \times i}$ *is given by*

$$Z = Y L_i^T, \tag{3.20}$$

*with*

$$L_i = \frac{\tau_m}{4} \begin{bmatrix} 2 & 0 & -1 & & & \\ & \frac{1}{2} & 0 & -\frac{1}{2} & & \\ & & \frac{1}{3} & 0 & \ddots & \\ & & & \frac{1}{4} & \ddots & -\frac{1}{i-2} \\ & & & & \ddots & 0 \\ & & & & & \frac{1}{i} \end{bmatrix} \in \mathbb{R}^{i \times i},$$

*and* $\hat{x} \in \mathbb{C}^{n \times 1}$ *satisfies*

$$\hat{x} = R_0^{-1} \left( \sum_{i=0}^{k-1} y_i - A_0 \sum_{i=0}^{k-1} z_i - \sum_{j=1}^{m} A_j \left( \sum_{i=0}^{k-1} T_{i+1}(1 - 2\frac{\tau_j}{\tau_m}) z_i \right) \right). \tag{3.21}$$

Theorem 3.4 states in a precise way that a vector having all but the $in$ first elements equal to zero is mapped by $G_N$ on a vector having all but the $(i+1)n$ first elements equal to zero. This property is a consequence of the *companion-like* structure of (3.10).

If we assume that the vector $b$ in (3.19) has the structure

$$b = [x_0^T \ 0 \ \cdots 0]^T, \quad x_0 \in \mathbb{C}^{n \times 1}, \tag{3.22}$$

then the vectors $G_N b, \ldots, G_N^{k-1} b$ only have their first $2n$, $3n$, $\ldots$, $kn$ elements different from zero. The Arnoldi algorithm builds the Krylov sequence vector by vector, where, in addition, the vectors are orthogonalized. In step $i$, the orthogonalization is a linear combination of the $(i+1)$st vector and the previously computed $i$ vectors. Hence, the orthogonalization at the $i$th iteration does not change the general structure of the $(i+1)$st vector. A dynamic implementation of Arnoldi to construct a basis of (3.19) with the starting vector (3.22), where this property is exploited, is described in Algorithm 1. In the description we use notation common for Arnoldi iterations: we let $\underline{\mathcal{H}}_i \in \mathbb{C}^{(i+1) \times i}$ denote the dynamically constructed rectangular Hessenberg matrix and $\mathcal{H}_i \in \mathbb{C}^{i \times i}$ the corresponding $i \times i$ upper part. To simplify the notation we will further denote the Krylov space (3.19) with starting vector (3.22) by $\mathcal{K}_k(G_N, x_0)$.

REMARK 3.5. *The construction of Algorithm 1 and, in particular, the Hessenberg matrix* $\mathcal{H}_k$, *do* not *depend on the value of* $N$. *The only constraint is that* $N \geq k$. *By taking the limit* $N \to \infty$ *and by taking into account that* $G_N$ *and* $A_N^{-1}$ *are*

**Algorithm 1** Dynamic construction of Krylov space

**Require:** $k \geq 1$, $x_0 \in \mathbb{C}^{n\times 1}$.

1: Let $v_1 = x_0/\|x_0\|_2$, $V_1 = v_1$, $\underline{\mathcal{H}}_0 =$ empty matrix
2: Factorize $R_0 = \sum_{i=0}^{m} A_i$
3: **for** $i = 1, 2, \ldots, k$ **do**
4: Let $\mathrm{vec}(Y) = v_i$
5: Compute $Z$ according to (3.20) with sparse $L_i$
6: Compute $\hat{x}$ according to (3.21) using the factorization of the inverse computed in Step 2
7: Expand $V_i$ with one block row (zeros)
8: Let $w_i := \mathrm{vec}(\hat{x}, Z)$, compute $h_i = V_i^* w_i$ and then $\hat{w}_i = w_i - V_i h_i$
9: Compute $\beta_i = \|\hat{w}_i\|_2$ and let $v_{i+1} = \hat{w}_i/\beta_i$
10: Let $\underline{\mathcal{H}}_i = \begin{bmatrix} \underline{\mathcal{H}}_{i-1} & h_i \\ 0 & \beta_i \end{bmatrix} \in \mathbb{C}^{(i+1)\times i}$
11: Expand $V_i$ into $V_{i+1} = [V_i, v_{i+1}]$
12: **end for**

**Output:** $\mathcal{H}_k$, $\underline{\mathcal{H}}_k$ $V_k$, $V_{k+1}$ $v_{k+1}$,
basis $\mathcal{K}_k(G_N, x_0)$, with $N \geq k$, by extending $V_k$ with $(N+1-k)n$ zero rows.

*similar matrices, the algorithm can be interpreted as an Arnoldi algorithm applied to the* infinite-dimensional *operator $\mathcal{A}^{-1}$, with $\mathcal{A}$ defined in §2.1. This connection is formalized and proven in [15, Section 4].*

REMARK 3.6. *Because the rightmost characteristic roots of (1.1), which correspond to the eigenvalues of $\mathcal{A}$, are typically among the smallest characteristic roots (this is apparent in Figure 2.1, see also [20, Chapter 2] for a detailed analysis of spectral properties of time-delay systems), an application of Algorithm 1, followed by computing the Ritz values (the eigenvalues of $\mathcal{H}_k$), are very efficient for computing the rightmost characteristic roots of the time-delays system (1.1). We refer to [15] for a detailed analysis.*

**3.3. Projection and moment matching properties.** We now arrive at the derivation of an approximation of $\gamma_N(s)$, defined by (2.8) or, equivalently, (3.9), having a prescribed order $k$. An approach to do so consists of constructing the Krylov space $\mathcal{K}_k(G_N, x_0)$ by Algorithm 1 and projecting the matrices $F_N, G_N, H_N$, defined in Theorem 3.2, on this Krylov space. Assuming $k \leq N$, an orthogonal projection on $\mathcal{K}_k(G_N, x_0)$ yields the following approximation of $\gamma_N(s)$:

$$\gamma^{(k)}(s) := F^{(k)} \ (sG^{(k)} - I)^{-1} \ H^{(k)} + D, \tag{3.23}$$

where

$$\begin{array}{rcl} F^{(k)} & = & F_{k-1} \ V_k, \\ G^{(k)} & = & \mathcal{H}_k, \\ H^{(k)} & = & V_k^T \ H_{k-1}, \end{array} \tag{3.24}$$

and the matrices

$$V_k = [v_1 \ \cdots \ v_k] \in \mathbb{R}^{kn\times k}, \mathcal{H}_k \in \mathbb{R}^{k\times k}$$

refer to the output of Algorithm 1. It is important to note that the matrices $F^{(k)}$ and $H^{(k)}$ are *submatrices* of $F^{(k+1)}$ and $H^{(k+1)}$. Therefore, they can be constructed in a

dynamic way when doing iterations of Algorithm 1, as is the case with the Hessenberg matrix $\mathcal{H}_k$.

With a particular choice of the vector $x_0$, the transfer function (3.23) satisfies the following moment matching property with the (original) transfer function (1.2) of the time-delay system (1.1).

THEOREM 3.7. *Let $N, k \in \mathbb{N}$ with $N \geq k \geq 2$, and let $V_k \in \mathbb{R}^{kn \times k}$. Assume that the columns of the matrix $V_k$, possibly extended with zero rows, form an orthogonal basis of $\mathcal{K}_k(G_N, R_0^{-1}B)$. Then the transfer function (3.23) satisfies*

$$\left.\frac{d^i \gamma^{(k)}(s)}{ds^i}\right|_{s=0} = \left.\frac{d^i \gamma(s)}{ds^i}\right|_{s=0}, \quad i = 0, \ldots, k-2 \tag{3.25}$$

*and*

$$\left.\frac{d^i \gamma^{(k)}(s^{-1})}{ds^i}\right|_{s=0} = \left.\frac{d^i \gamma(s^{-1})}{ds^i}\right|_{s=0}, \quad i = 0, 1. \tag{3.26}$$

*Proof.* The proof is performed in two steps. First, we observe that

$$G_N^{-1} H_N = \left[(R_0^{-1}B)^T \ 0 \ \cdots \ 0\right]^T,$$

from which we conclude

$$\mathcal{K}_k(G_N, R_0^{-1}B) = \operatorname{span}\left\{G_N^{-1}H_N, H_N, G_N H_N, \ldots, G^{k-2}H_N\right\}. \tag{3.27}$$

It follows that the transfer function $\gamma^{(k)}(s)$ matches $k-1$ moments at zero and two at infinity with the transfer function $\gamma_N(s)$. Second, this moment matching property carries over to the transfer function $\gamma(s)$ of the delay equation by Theorem 2.1. ∎

The principle behind the proof of Theorem 3.7, along with an overview of the results obtained so far, are shown in Figure 3.1.

REMARK 3.8. *The derivation of the matrices $(F_N, G_N, H_N)$ in the proof of Theorem 3.2 was based on the choice*

$$\begin{aligned} H_N &= G_N^2 (S_N^{-1} \otimes I) B_N, \\ F_N &= C_N (S_N \otimes I) G_N^{-1}. \end{aligned}$$

*Another realization of the transfer function $\gamma_N(s)$ can be obtained by taking*

$$\begin{aligned} H_N &= G_N^j (S_N^{-1} \otimes I) B_N, \\ F_N &= C_N (S_N \otimes I) G_N^{1-j}, \end{aligned}$$

*for any $j \in \mathbb{Z}$. With the choice $j = 2$ the matrix $H_N$ has the structure that leads to (3.27).*

Observe that in the reduced model (3.23)-(3.24) *none* of the system matrices depend on $N$. This is a consequence of Proposition 3.3 and the special structure of the starting vector of the Arnoldi process. Since (3.23) corresponds to a projection on a $k$-dimensional subspace of (2.7) for any $N \geq k$, we can take a limit argument ($N \to \infty$), and interpret (3.23) as the transfer function of a projection of the infinite-dimensional linear system (2.1). This is consistent with the moment matching property described in Theorem 3.7. In the light of this comment, the discretized system (2.7) only serves

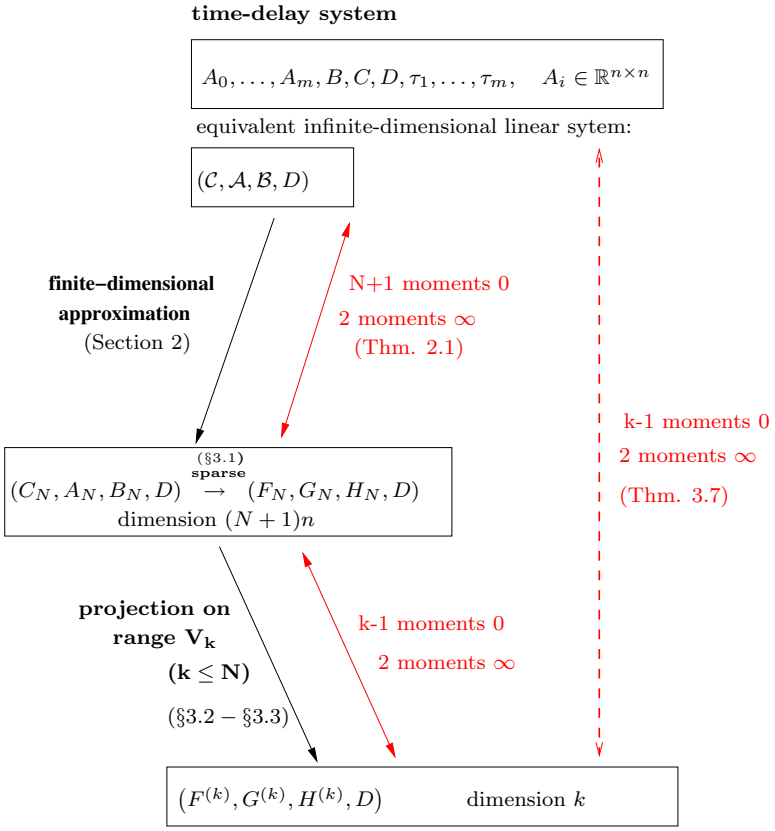



FIG. 3.1. *Overview of Sections 2-3. The results on moment matching are shown in red.*

as an intermediate step in the technical derivation of the moment matching property, see Figure 3.1.

REMARK 3.9. The matching of $k-1$ moments at zero is due to the fact that the space on which we project, contains an appropriately defined right Krylov space of $G_N$. It is well known that twice as many moments can be matched by constructing both a right and a left Krylov space (bi-orthogonal bases can be found by, e.g., the two-sided Arnoldi algorithm [7]) and an oblique projection. This approach has not been followed because of the following reasons.

1. The efficient and dynamic construction of the (right) Krylov space, described in Algorithm 1, relies on the companion-like structure of the matrix $G_N$, where the system-dependent information is in the first block *row*. This construction does not carry over to the left Krylov space. By taking the transpose of the transfer function the role of the left and right Krylov spaces are interchanged, but a simultaneous dynamic construction of a left and a right Krylov space as in Algorithm 1 is not possible.
2. Among the main arguments for deriving the reduced model starting from a spectral discretization of the time-delay system (Section 2) are the excellent properties of a spectral discretization in approximating the rightmost characteristic roots [5] and the interpretation of Algorithm 1 as an Arnoldi algorithm applied to the inverse of the infinite-dimensional operator $\mathcal{A}$, whose eigenvalues correspond the

characteristic roots [15].

To conclude the section, we describe the overall model reduction approach in Algorithm 2.

**Algorithm 2** Derivation reduced model of dimension $k$ ($k \geq 2$).

1: Apply Algorithm 1 with $x_0 = R_0^{-1}B$ and construct $G^{(k)} = \mathcal{H}_k$. At the same time dynamically construct $F^{(k)}$ and $H^{(k)}$, defined in (3.24).

**Output:** Matrices $(F^{(k)}, G^{(k)}, H^{(k)}, D)$ of the reduced model;
$\gamma^{(k)}(s) = F^{(k)}(sG^{(k)} - I)^{-1}H^{(k)} + D$.

**3.4. Note on the computational complexity.** Many model reduction applications involve the reduction of a very large original system. We will now describe the complexity of Algorithm 2 for the case where $n$ is large and $k$ is small or moderate.

| | Algorithm 2 | delay free system |
|---|---|---|
| nof. backward solves | $k+1$ | $k-1$ |
| nof. matrix vector products | $O(k)$ | 1 |
| nof. scalar products (orthogonalization) | $O(k^3)$ | $O(k^2)$ |

TABLE 3.1

*Operation count and computational complexity for Algorithm 2 and the standard Krylov model reduction technique for delay free systems.*

The asymptotic complexity for the possibly dominating parts of Algorithm 2 are given in the first column of Table 3.1. We use the terminology *backward solve* to refer to the solving of the corresponding linear system of equations, in this case solving the linear system in (3.21). To ease the comparison we have given the counts (for the first column) in terms of the number of operations associated with vectors of length $n$. That is, computing the product $V_i^* w_i$ in Step 8 of Algorithm 1 has essentially the same computational effort as computing $i^2$ scalar products for vectors of length $n$. Moreover, in each iteration of Algorithm 1 we need $m+1$ matrix vector operations to compute $\hat{x}$ in (3.21). Hence in total, we need $O(k)$ matrix vector multiplications and $O(k^3)$ scalar products.

Consider a standard dynamical (delay-free) system given in the form (2.7). Such a system can be reduced with a standard Krylov moment matching technique. Note that in order to match 2 moments at infinity and $k-1$ moments at $s=0$ we need 1 matrix vector product, $k-1$ backward solves and $O(k^2)$ scalar products. We give the corresponding counts in the last column of Table 3.1.

We will now see that several advantages in terms of complexity of Algorithm 2 are apparent from Table 3.1.

Time-delay systems are generally considered more difficult to analyze than delay-free systems, as they are in fact infinite-dimensional systems. In a modeling situation one might be faced with the choice of incorporating the delay or settling for a cruder model by neglecting the delay entirely. From Table 3.1 we conclude that the difference between carrying out a model reduction algorithm for a time-delay system of dimension $n$ and doing a model reduction on a (cruder) model without delay (also of dimension $n$) is not tremendous. In fact, for sufficiently large $n$, the computational effort for the matrix vector product is typically not dominating and the quotient is

essentially $k$ (which is considered small or moderate). Hence, taking a delay into account in a model does not introduce a big additional computational cost.

From the counts in Table 3.1 we can also compare Algorithm 2 with a direct model-reduction approach, which we refer to as `discretize+reduce`. The discretized system (2.7), where we choose $N = k - 2$, is of dimension $n(k-1)$. Since it is a standard dynamical system with moments coinciding with the original system (1.1), more precisely $k-1$ at zero and two at infinity, an intuitive approach to model reduction of (1.1) would be to use a standard Padé via Krylov technique on (2.7) such that these matching moments are carried over to the reduced model. In this approach we would need $k-1$ backward solves for a $(k-1)n \times (k-1)n$ matrix. The matrix $A_N$ is not sparse even if the original problem is sparse. This should be compared with $k+1$ solves with a $n \times n$ matrix for Algorithm 2. Clearly, $k+1$ backward solves with $n \times n$ matrix is for sufficiently large $n$ is expected to be much faster than $k-1$ backward solves for $(k-1)n \times (k-1)n$ matrix. The computational effort for the scalar products is also smaller for Algorithm 2. Note that with the `discretize+reduce` approach, one would compute $k^2$ scalar products with vectors of length $(k-1)n$, which is essentially equivalent to $k^2(k-1)$ scalar product with vectors of length $n$. Algorithm 1 involves less scalar products, due to the structure of the basis vectors of the subspace, with only $n, 2n, \ldots, kn$, elements different from zero.

A major advantage of the proposed method is that Algorithm 2 is an iterative procedure. This allows inspection of error during the iteration and we have the possibility to continue the iteration if the accuracy is deemed insufficient. This dynamic feature is not present in the `discretize+reduce` approach, where the number of discretization points has to be chosen before starting the method.

**4. Application.** We study a heated rod which is cooled using delayed feedback. The algorithm of this paper can be used in such a study by computing an accurate dynamical system of small dimension which does not involve a delay. This is here achieved by first discretizing a heat equation in space and applying the model reduction algorithm to the discretized problem, which is a time-delay system of the form (1.1).

The physical model of the heated rod which we consider in this example is given by the partial differential equation

$$\frac{\partial v(x,t)}{\partial t} = \frac{\partial^2 v(x,t)}{\partial x^2} + a_0(x)v(x,t) + a_1(x)v(\pi - x, t-1). \tag{4.1}$$

with $a_0(x) = -2\sin(x)$, $a_1(x) = 2\sin(x)$ and $v(0,t) = v(\pi,t) = 0$, also used in [15]. The equation is a variant of [6, Example C] and can be interpreted as the heat equation corresponding to a rod with a distributed heating source and a non-local weighted delayed feedback. We discretize the differential equation (4.1) in space such that the corresponding time-delay system is of dimension $n$ and fit it with the output matrix $C = (1, 1, \ldots, 1)/\|(1, 1 \ldots, 1)\|_2$, i.e., the output is the average temperature of the rod. The model reduction algorithm is applied to the system for two different choices of the input matrix $B$,

i) $B = B_1 := C^T$, and,

ii) $B = B_2 := e_{n/5}$, i.e., the control $Bu(t)$ is localized at position $x = \pi/5$.

The corresponding time-delay systems of dimension $n = 100$ are now reduced by Algorithm 2 with $k = 20$. As expected from the theory, the reduced model is accurate both in terms of frequency response and in terms of approximation of the

characteristic roots of the system. We illustrate this with figures, which should be interpreted as follows.

The frequency response and the point-wise error in the frequency response for both choices of $B$ are plotted in Figure 4.1. We clearly see that the approximation is good at $\omega = 0$ where many moments are matched and also decreasing for large frequencies.

Because the model reduction approach has a connection with a spectral discretization of the linear infinite-dimensional system (2.1) and Algorithm 1 can be interpreted as an Arnoldi algorithm applied to the operator $\mathcal{A}^{-1}$, we also expect the relevant eigenvalues of the reduced system to be good approximations of the characteristic roots of the original time-delay system. In Figure 4.2 we observe that many roots close to the origin are well approximated in the reduced model. Note that for $B = B_1$ some characteristic roots are not in the reduced model, e.g., $s \approx -1.3 \pm 2i$. This property is consistent with the fact that Algorithm 2 corresponds to a projection on a Krylov space. The roots that are not captured correspond to eigenvalues for which the left eigenfunctions are orthogonal to $B$ (as an illustration, the sum of the elements of the eigenvector corresponding to $s \approx -1.3 - 2i$ vanish due to the symmetry of the problem, see Figure 4.3). In a control setting these characteristic roots are so-called uncontrollable modes. The fact that they are not present in the reduced model can be seen as a positive feature, because they do not appear as poles in the original transfer function. In Figure 4.4 we illustrate the convergence of the first eigenvalues of $G^{(k)}$. The convergence is exponential, which is in accordance with an Arnoldi method.

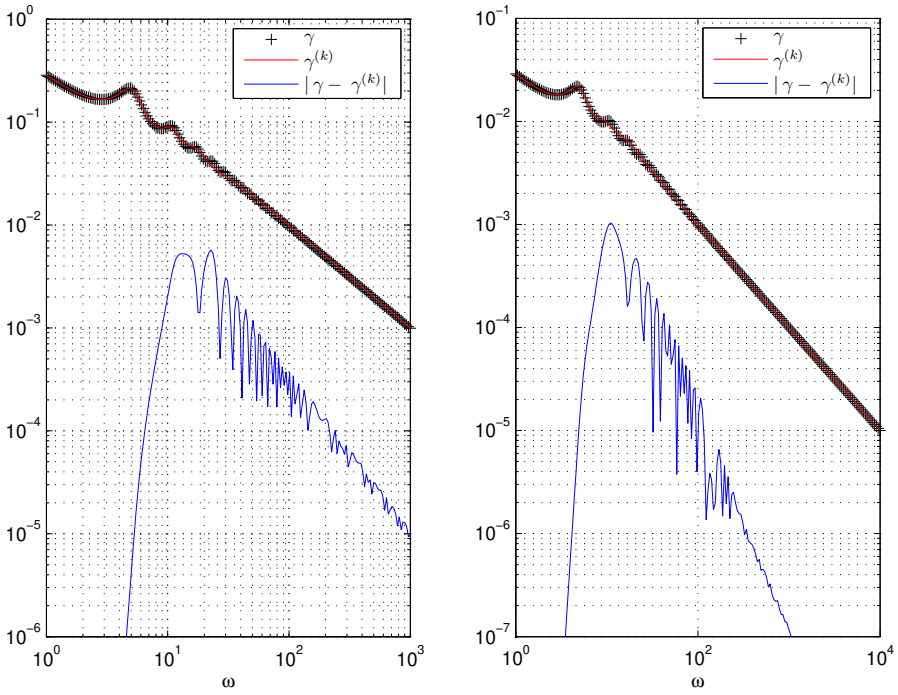
Finally we note that the rightmost eigenvalues of $(G^{(k)})^{-1}$ coincide with the rightmost characteristic roots, see Figure 4.2. As a consequence, the reduced models preserve the stability of the system.

**5. Conclusions.** We proposed an approach for deriving reduced order models for infinite-dimensional time-delay systems. It relies on a dynamic construction of a Krylov space and a projection on this space. Because the matrices involved stem from a spectral discretization, the method has the property of *well capturing the rightmost characteristic roots*. The construction is dynamic in the sense that the number of discretizaton points, $N + 1$, does *not* need to be chosen and the construction can always be resumed if the accuracy of the reduced model turns out to be insufficient. This property further allows to interpret the reduced model directly as a projection of the infinite-dimensional system (2.1). In particular, the matrix $G^{(k)} = \mathcal{H}_k$, with $k$ the number of iterations, can be interpreted as the result of $k$ Arnoldi iterations applied to the inverse of the infinite-dimensional operator $\mathcal{A}$, whose eigenvalues correspond to the characteristic roots.

The model reduction approach also results in a *moment matching property*: the transfer function of the reduced model matches $k - 1$ moments at zero and two moments at infinity with the transfer function of the time-delay system.

In all our experiments the rightmost eigenvalues of the reduced model corresponded to the rightmost characteristic roots (provided a controllability condition is satisfied), or, equivalently, the eigenvalues of the reduced model that had not yet converged to characteristic roots were located to the left of the converged eigenvalues. This favorable property, which carries over from the spectral discretization, makes the reduced model suitable in the context of control design. A detailed analysis of this phenomenon, including the connection with the position of the pseudospectra contours [19, 26] is beyond the scope of this paper and topic of further research.

As a byproduct we showed that a spectral discretization of an infinite-dimensional




i) $B = B_1$ ii) $B = B_2$

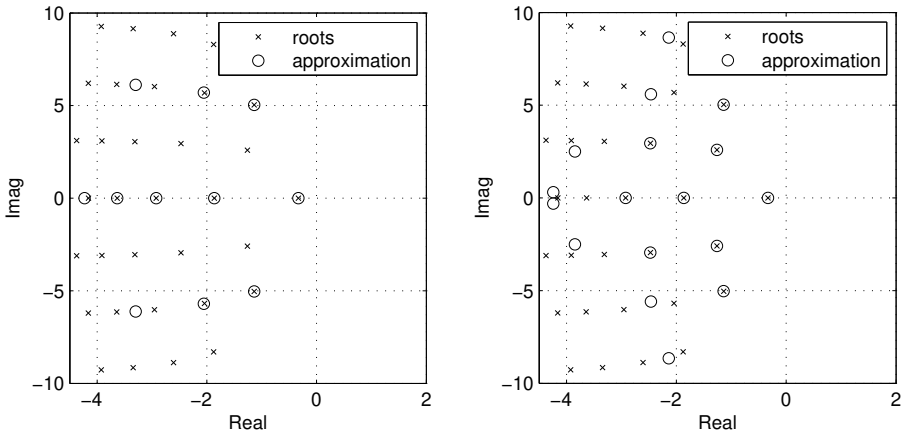
FIG. 4.1. *Frequency response and error for the example in Section 4. The reduced model is of dimension* $k = 20$

time-delay systems results in a moment matching property (Theorem 2.1), *independently* of the choice of the grid points. This illustrates the importance of imposing additional conditions. In the article the grid points were chosen in such a way that the right most characteristic roots were well approximated (spectral convergence of the individual eigenvalues) and, in addition, a reformulation of the discretized system in a sparse, companion-like representation became possible (Theorem 3.2). The latter was at the basis of the dynamic construction of the Krylov space. The choice of the grid points also led to the special structure of the matrix $H_N$, which allowed to apply Algorithm 1 in the context of a Krylov-Padé reduction approach.

The approach of the paper can be extended to systems with multiple inputs and multiple outputs provided that 'block versions' of the algorithms are used (in particular, block Arnoldi in Algorithm 1).

**Acknowledgements.** This article present results of the Belgian Programme on Interuniversity Poles of Attraction, initiated by the Belgian State, Prime Minister's Office for Science, Technology and Culture, the Optimization in Engineering Centre OPTEC, and the project STRT1-09/33 of the K.U.Leuven Research Foundation.

REFERENCES




i) $B = B_1$ ii) $B = B_2$

FIG. 4.2. *Characteristic roots and approximations (eigenvalues of* $\left(G^{(k)}\right)^{-1}$*) for the example in Section 4.*

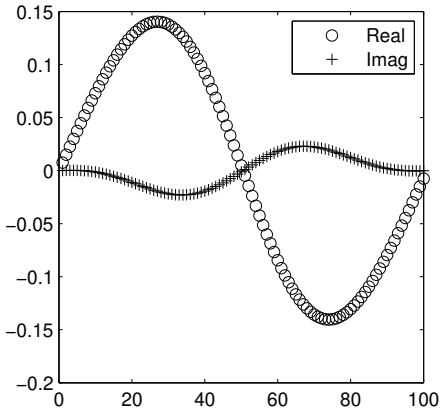



FIG. 4.3. *The left eigenvector corresponding to eigenvalue* $s \approx -1.3 - 2i$*.*

[1] A.C. Antoulas, C.A. Beattie, and S. Gugercin. *Interpolatory model reduction of large-scale dynamical systems*, page 56. Springer-Verlag, 2010.

[2] Z. Bai and R. W. Freund. A partial Padé-via-Lanczos method for reduced-order modeling. *Linear Alg. Appl.*, 332–334:139–164, 2001.

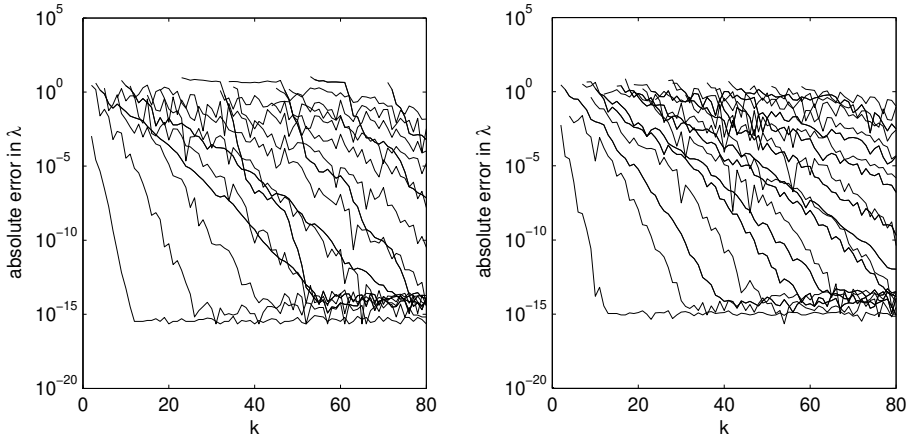
[3] Z. Bai and Y. Su. Dimension reduction of second-order dynamical systems via a second-order Arnoldi method. *SIAM J. Matrix Anal. and Applic.*, 26(5):1692–1709, 2005.

[4] C. Beattie and S. Gugercin. Interpolatory projection methods for structure-preserving model reduction. *Syst. Control Lett.*, 58(3):225–232, 2009.

[5] D. Breda, S. Maset, and R. Vermiglio. Pseudospectral differencing methods for characteristic roots of delay differential equations. *SIAM J. Sci. Comput.*, 27(2):482–495, 2005.

[6] D. Breda, S. Maset, and R. Vermiglio. Numerical approximation of characteristic values of partial retarded functional differential equations. *Numer. Math.*, 113(2):181–242, 2009. doi:10.1007/s00211-009-0233-7.

[7] J. Cullum and T. Zhang. Two-sided Arnoldi and nonsymmetric Lanczos algorithms. *SIAM J. Matrix Anal. Appl.*, 24(2):303–319, 2002.




i) $B = B_1$ ii) $B = B_2$

FIG. 4.4. *The exponential convergence of the first characteristic roots.*


[8] R. F. Curtain and H. Zwart. *An introduction to infinite-dimensional linear systems theory*. Springer-Verlag, NY, 1995.

[9] P. Feldman and R. W. Freund. Efficient linear circuit analysis by Padé approximation via the Lanczos process. *IEEE Trans. Computer-Aided Design*, CAD-14:639–649, 1995.

[10] R. W. Freund. Subspaces associated with higher-order linear dynamical systems. *BIT*, 45:495–516, 2005.

[11] K. Gallivan, E. Grimme, and P. Van Dooren. A rational Lanczos algorithm for model reduction. *Numerical Algorithms*, 12:33–63, 1996.

[12] E. Grimme, D. Sorensen, and P. Van Dooren. Model reduction of state space systems via an implicitly restarted Lanczos method. *Numerical Algorithms*, 12:1–31, 1996.

[13] S. Gugercin. An iterative SVD-Krylov based method for model reduction of large-scale dynamical systems. *Linear Alg. Appl.*, 428:1964–1986, 2008.

[14] S. Gumussoy and W. Michiels. A predictor-corrector type algorithm for the pseudospectral abscissa computation of time-delay systems. *Automatica*, 46(4):657–664, 2010.

[15] E. Jarlebring, K. Meerbergen, and W. Michiels. A Krylov method for the delay eigenvalue problem. Technical Report TW558, K.U. Leuven, 2010. submitted.

[16] P. Mäkilä and J. Partington. Laguerre and Kautz shift approximations of delay systems. *Int. J. Control*, 72(10):932–946, 1999.

[17] P. Mäkilä and J. Partington. Shift operator induced approximations of delay systems. *SIAM J. Control Optimization*, 37(6):1897–1912, 1999.

[18] K. Meerbergen. The quadratic Arnoldi method for the solution of the quadratic eigenvalue problem. *SIAM J. Matrix Anal. Appl.*, 30(4):1463–1482, 2008.

[19] W. Michiels, K. Green, T. Wagenknecht, and S.-I. Niculescu. Pseudospectra and stability radii for analytic matrix functions with application to time-delay systems. *Lin. Alg. Appl.*, 418(1):315–335., 2006.

[20] W. Michiels and S.-I. Niculescu. *Stability and Stabilization of Time-Delay Systems: An Eigenvalue-Based Approach*. Advances in Design and Control 12. SIAM Publications, Philadelphia, 2007.

[21] J. Partington. Model reduction of delay systems. In V. Blondel and A. Megretski, editors, *Unsolved Problems in Mathematical Systems and Control Theory*, pages 29–32. Princeton university press, 2004.

[22] J. Partington. Some frequency-domain approaches to the model reduction of delay systems. *Annual Reviews in Control*, 28(1):65–73, 2004.

[23] M. Rewieński and J. White. A trajectory piecewise-linear approach to model order reduction and fast simulation of nonlinear circuits and micromachined devices. In *ICCAD '01: Proceedings of the 2001 IEEE/ACM international conference on Computer-aided design*, pages 252–257, Piscataway, NJ, USA, 2001. IEEE Press.

[24] V. Simoncini and F. Perotti. On the numerical solution of $(\lambda^2 A+\lambda B+C)x = b$ and application

to structural dynamics. *SIAM J. on Sci. Comput.*, 23(6):1875–1897, 2002.
[25] L. N. Trefethen. *Spectral Methods in MATLAB*. SIAM Publications, Philadelphia, 2000.
[26] T. Wagenknecht, W. Michiels, and K. Green. Structured pseudospectra for nonlinear eigenvalue problems. *J. Comput. Appl. Math.*, 211:245–259, 2008. doi: 10.1016/j.cam.2006.12.00.