



Data-driven model reduction by moment matching for linear and nonlinear systems[☆]



Giordano Scarciootti^a, Alessandro Astolfi^{a,b}

^a Department of Electrical and Electronic Engineering, Imperial College London, London, SW7 2AZ, UK

^b DICII, University of Rome “Tor Vergata”, Via del Politecnico 1, 00133 Rome, Italy

ARTICLE INFO

Article history:

Received 9 November 2015

Received in revised form

25 October 2016

Accepted 6 January 2017

Available online 6 March 2017

Keywords:

Model reduction

System identification

Model reduction from data

Moment matching

ABSTRACT

Theory and methods to obtain reduced order models by moment matching from input/output data are presented. Algorithms for the estimation of the moments of linear and nonlinear systems are proposed. The estimates are exploited to construct families of reduced order models. These models asymptotically match the moments of the unknown system to be reduced. Conditions to enforce additional properties, e.g. matching with prescribed eigenvalues, upon the reduced order model are provided and discussed. The computational complexity of the algorithms is analyzed and their use is illustrated by two examples: we compute converging reduced order models for a linear system describing the model of a building and we provide, exploiting an approximation of the moment, a nonlinear planar reduced order model for a nonlinear DC-to-DC converter.

© 2017 The Author(s). Published by Elsevier Ltd.

This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The availability of mathematical models is essential for the analysis, control and design of modern technological devices. As the computational power has advanced, the complexity of these mathematical descriptions has increased. This has kept the computational needs at the top or over the available possibilities (Åström & Kumar, 2014). A solution to this problem is represented by model reduction which consists in finding a simplified mathematical model which maintains some key properties of the original description.

In the linear framework several techniques have been developed by the systems and control community (Antoulas, 2005). These techniques can be divided into two main groups: approximation methods based on singular value decomposition (SVD) and approximation methods based on Krylov projectors, also known as

moment matching methods. The use of the notion of Hankel operators (Adamjan, Arov, & Krein, 1971; Glover, 1984; Safonov, Chiang, & Limebeer, 1990) and the theory of balanced realizations (Gray & Mesko, 1997; Lall & Beck, 2003; Meyer, 1990; Moore, 1981) belong to the first group, whereas the use of interpolation theory and the notion of projections (Antoulas, Ball, Kang, & Willems, 1990; Beattie & Gugercin, 2008; Byrnes, Lindquist, & Georgiou, 2001; Byrnes, Lindquist, Gusev, & Matveev, 1995; Gallivan, Vandendorpe, & Van Dooren, 2004, 2006; Georgiou, 1999; Gugercin, Antoulas, & Beattie, 2008; Kimura, 1986) belong to the latter. In the nonlinear framework, a continuous effort has been devoted to extend the linear techniques or develop original approaches. Results for special classes of nonlinear systems, such as bilinear systems and Hamiltonian systems, have been presented in Al-Baiyat, Bettayeb, and Al-Saggaf (1994), Fujimoto (2008), Lall, Krysl, and Marsden (2003) and Soberg, Fujimoto, and Glad (2007). Nonlinear “enhancements” of the balancing techniques have been proposed in Besselink, van de Wouw, and Nijmeijer (2012, 2013), Besselink, van de Wouw, Scherpen, and Nijmeijer (2014), Fujimoto and Scherpen (2010), Fujimoto and Tsubakino (2008), Gray and Mesko (1997, 1999), Gray and Scherpen (2005), Lall, Marsden, and Glavaski (2002), Scherpen and Gray (2000, 2002) and Willcox and Peraire (2002). Reduction of nonlinear systems around invariant manifolds represents another direction of investigation, see e.g. Gray and Verriest (2006) and Verriest and Gray (1998). Model reduction methods based on proper orthogonal decomposition (POD) have been developed for linear and nonlinear systems, see e.g. Astrid, Weiland,

[☆] This work was supported in part by Imperial College London under the Junior Research Fellowship Scheme (number EESC-F38010) and in part by the EPSRC grant STABLE-NET (number EP/L014343/1). The material in this paper was partially presented at the 14th annual European Control Conference, July 15–17, 2015, Linz, Austria and at the 54th IEEE Conference on Decision and Control, December 15–18, 2015, Osaka, Japan. This paper was recommended for publication in revised form by Associate Editor Lorenzo Marconi under the direction of Editor Andrew R. Teel.

E-mail addresses: g.scarciootti@ic.ac.uk (G. Scarciootti), a.astolfi@ic.ac.uk (A. Astolfi).

Willcox, and Backx (2008), Hinze and Volkwein (2005), Kunisch and Volkwein (1999, 2008), Rowley, Colonius, and Murray (2004) and Willcox and Peraire (2002). A nonlinear enhancement of the moment matching approach has eluded researchers until recently. In Gallivan et al. (2004, 2006), a characterization of the moments in terms of the solution of a Sylvester equation has been given. Exploiting this relation, in Astolfi (2007, 2010) a new interpretation of the moment matching approach, based on a steady-state description of moment, has been proposed. This has led to further developments in the model reduction field, e.g. the extension of the model reduction theory to linear and nonlinear time-delay systems (Scarcitti & Astolfi, 2016b), see also Ionescu and Astolfi (2013b), Ionescu, Astolfi, and Colaneri (2014), Scarcitti (2015) and Scarcitti and Astolfi (2015a, 2016a).

When dealing with model reduction, one usually starts with a high-dimensional model to be reduced. In fact, most of the methods assume the knowledge of a state-space model of the system to be reduced. However, in practice this model is not always available. Among the moment matching methods, a data-driven approach to reduce linear systems has been proposed under the name of Loewner framework (Mayo & Antoulas, 2007). This method constructs reduced order models by using matrices composed of frequency-domain measurements, which makes intrinsically difficult to extend the method to nonlinear systems. In this paper, inspired by the learning algorithm given in Bian, Jiang, and Jiang (2014) and Jiang and Jiang (2012, 2014) to solve a model-free adaptive dynamic programming problem (see also the references therein, e.g. Baird, 1994; Vrabie, Pastravanu, Abu-Khalaf, & Lewis, 2009), we propose *time-domain* data-driven on-line algorithms for the model reduction of linear and nonlinear systems. Collecting *time-snapshots* of the input and output of the system at a given sequence of time instants t_k , two algorithms to define families of reduced order models (in the framework introduced in Astolfi, 2010) at each instant of the iteration t_k are devised. The reduced order models asymptotically match the moments of the unknown system to be reduced. These algorithms have several advantages with respect to an identification plus reduction technique: there is no need to identify the system, which is expensive both in terms of computational power and storage memory; since the reduced order model matches the moments of the unknown system, it is not just the result of a low-order identification but it actually retains some properties of the larger system; since the proposed algorithm determines directly the moment of a nonlinear system from the input and output data, it does not involve the computation of the solution of a partial differential equation, which is usually a difficult task; in addition, for linear systems the capability of determining reduced order models from data matrices (of order proportional to ν) is computationally efficient when compared with model reduction obtained by manipulating the system matrices (of order $n \gg \nu$). Thus, the method is of computational value, for both linear and nonlinear systems, even when the system to be reduced is known.

This work is at the intersection of large and active research areas, such as model reduction and system identification. For instance, the use of time-snapshots and data matrices is reminiscent of the methods of POD, see e.g. Lorenz (1956) and Noack, Afanasiev, Morzynski, Tadmor, and Thiele (2003) and of subspace identification, see e.g. van Overschee and de Moor (1996) and Verhaegen and Verdult (2007). Moreover, while the Loewner framework is a frequency-domain data-driven method proposed in model reduction, various time-domain data-driven techniques have been presented in system identification, such as the eigensystem realization algorithm, see e.g. Cooper (1999), Houtzager, van Wingerden, and Verhaegen (2012), Majji, Juang, and Junkins (2010) and Rebolho, Belo, and Marques (2014), and the dynamic mode decomposition, see Hemati, Williams, and Rowley (2014). Hence, this paper can be

seen as part of the model reduction literature or of the system identification literature. However, since the results of the paper have strong connections with the notion of moment, we present the results using the terminology of model reduction.

The rest of the paper is organized as follows. In Section 2.1 (3.1) the definition of moment and the model reduction techniques, as developed in Astolfi (2010), are recalled for linear (nonlinear) systems. In Section 2.2 we give a preliminary analysis to compute on-line estimates of the moments of a linear system. In Section 2.3 (3.2) approximations which converge asymptotically to the moments of the linear (nonlinear) system are given. Therein, a discussion of the computational complexity associated with the evaluation of these approximations is presented, a recursive least-square formula is given and a moment estimation algorithm is provided. In Section 2.4 (3.3) we give a family of reduced order models for linear (nonlinear) systems. In Section 2.5 we discuss how several properties, e.g. matching with prescribed eigenvalues or zeros, can be enforced in the present scenario. In Section 2.6 a linear reduced order model describing the dynamics of a building is estimated with the method proposed in the paper. In Section 3.4 a nonlinear reduced order model constructed using an approximation of the moment of a DC-to-DC converter provides a further example. Finally Section 4 contains some concluding remarks.

Preliminary versions of this paper have been published in Scarcitti and Astolfi (2015b,c). The additional contributions of the present paper are as follows: the results are presented in a more formal and organized way (by means, e.g. of theorems and propositions) and all the results are directly proved; two recursive algorithms are given; properties of the exponentially converging models, such as matching with prescribed eigenvalues, are discussed; the reduction of the linear model describing the dynamics of a building provides a new example based on a physical system; the so-called “ $\mathcal{U}/\mathcal{Y}$ ” variation of the nonlinear algorithm is given. Finally, the example in Scarcitti and Astolfi (2015c) is extended providing a nonlinear planar reduced order model estimated with the presented techniques.

While this paper was under review, the proposed algorithm has been applied to the problem of model reduction of power systems (Scarcitti, 2017). Therein the authors provide an extensive testing of the linear algorithm. This suggests that the results of the present paper can be of relevance for applications in diverse research domains.

Notation. We use standard notation. $\mathbb{R}_{\geq 0}$ ($\mathbb{R}_{>0}$) denotes the set of non-negative (positive) real numbers; $\mathbb{C}_{<0}$ denotes the set of complex numbers with negative real part; $\mathbb{C}_0$ denotes the set of complex numbers with zero real part. The symbol I denotes the identity matrix and $\sigma(A)$ denotes the spectrum of the matrix $A \in \mathbb{R}^{n \times n}$. The symbol $\otimes$ indicates the Kronecker product and $\|A\|$ indicates the induced Euclidean matrix norm. The symbol ϵ_k indicates a vector with the k th element equal to 1 and with all the other elements equal to 0. The vectorization of a matrix $A \in \mathbb{R}^{n \times m}$, denoted by $\text{vec}(A)$, is the $nm \times 1$ vector obtained by stacking the columns of the matrix A one on top of the other, namely $\text{vec}(A) = [a_1^T, a_2^T, \dots, a_m^T]^T$, where $a_i \in \mathbb{R}^n$ is the i th column of A and the superscript $\top$ denotes the transposition operator.

2. Linear systems

2.1. Model reduction by moment matching – recalled

To render the paper self-contained in this section we recall the notion of moment for linear systems as presented in Astolfi (2010). Consider a linear, single-input, single-output, continuous-time, system described by the equations

$$\dot{x} = Ax + Bu, \quad y = Cx, \quad (1)$$

with $x(t) \in \mathbb{R}^n$, $u(t) \in \mathbb{R}$, $y(t) \in \mathbb{R}$, $A \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{n \times 1}$ and $C \in \mathbb{R}^{1 \times n}$. Let $W(s) = C(sI - A)^{-1}B$ be the associated transfer function and assume that (1) is minimal, i.e. controllable and observable.

Definition 1. Let $s_i \in \mathbb{C} \setminus \sigma(A)$. The 0-moment of system (1) at s_i is the complex number $\eta_0(s_i) = W(s_i)$. The k -moment of system (1) at s_i is the complex number $\eta_k(s_i) = \frac{(-1)^k}{k!} \left[\frac{d^k}{ds^k} W(s) \right]_{s=s_i}$, with $k \geq 1$ integer.

In Astolfi (2010), exploiting the observation that the moments of system (1) can be characterized in terms of the solution of a Sylvester equation (see Gallivan et al., 2004, 2006), it has been noted that the moments are in one-to-one relation with the well-defined steady-state output response of the interconnection between a signal generator and system (1). This interpretation of the notion of moment, which relies upon the center manifold theory, has the advantage that it can be extended to nonlinear systems and it is of particular interest for the aims of this paper.

Theorem 1 (Astolfi, 2010). Consider system (1) and suppose $s_i \in \mathbb{C} \setminus \sigma(A)$, for all $i = 1, \dots, \eta$. Consider a non-derogatory¹ matrix $S \in \mathbb{R}^{v \times v}$ with characteristic polynomial $p(s) = \prod_{i=1}^{\eta} (s - s_i)^{k_i}$, where $v = \sum_{i=1}^{\eta} k_i$. Then there exists a one-to-one² relation between the moments $\eta_0(s_1), \dots, \eta_{k_1-1}(s_1), \dots, \eta_0(s_\eta), \dots, \eta_{k_\eta-1}(s_\eta)$ and

- the matrix $C\Pi$, where Π is the unique solution of the Sylvester equation

$$A\Pi + B L = \Pi S, \quad (2)$$

for any $L \in \mathbb{R}^{1 \times v}$ such that the pair (L, S) is observable;

- the steady-state response, provided $\sigma(A) \subset \mathbb{C}_{<0}$, of the output y of the interconnection of system (1) with the system

$$\dot{\omega} = S\omega, \quad u = L\omega, \quad (3)$$

for any L and $\omega(0)$ such that the triple $(L, S, \omega(0))$ is minimal.

Moreover, system (1) has a global invariant manifold described by $\mathcal{M} = \{(x, \omega) \in \mathbb{R}^{n+v} : x = \Pi\omega\}$. Hence, for all $t \in \mathbb{R}$,

$$x(t) = \Pi\omega(t) + e^{At}(x(0) - \Pi\omega(0)). \quad (4)$$

Finally, as shown in Astolfi (2010), the family of systems

$$\dot{\xi} = (S - GL)\xi + Gu, \quad \psi = C\Pi\xi, \quad (5)$$

with G any matrix such that $\sigma(S) \cap \sigma(S - GL) = \emptyset$, contains all the models of dimension v interpolating the moments of system (1) at the eigenvalues of S . Hence, we say that system (5) is a model of (1) at S .

2.2. A preliminary analysis

In this section we provide a preliminary analysis assuming to know the matrices A, B, C and the state $x(0)$ in Eq. (1). This analysis is used in the following section for the development of an estimation algorithm which, this time, does not use the matrices A, B, C and the state $x(0)$. To this end we make the following assumptions.

Assumption 1. The input u of system (1) is given by the signal generator (3), with S such that $\sigma(S) \subset \mathbb{C}_0$ with simple eigenvalues. In addition, assume that the triple $(L, S, \omega(0))$ is minimal.

¹ A matrix is non-derogatory if its characteristic and minimal polynomials coincide.

² The matrices A, B, C and the points s_i identify the moments. Then, given any observable pair (L, S) with $s_i \in \sigma(S)$, there exists an invertible matrix $T \in \mathbb{R}^{v \times v}$ such that the elements of the vector $CIT T^{-1}$ are equal to the moments.

Assumption 2. System (1) is asymptotically stable, i.e. $\sigma(A) \subset \mathbb{C}_{<0}$.

Assumption 1 has a series of implications. The hypothesis on the eigenvalues of S is reasonable since the contribution of the negative eigenvalues of S to the response of the system decays to zero. The minimality of the triple $(L, S, \omega(0))$ implies the observability of the pair (L, S) and the “controllability” of the pair $(S, \omega(0))$. This last condition, called *excitability* of the pair $(S, \omega(0))$, is a geometric characterization of the property that the signals generated by (3) are persistently exciting, see Åström and Wittenmark (1995), Padoan, Scarcioiti, and Astolfi (2016) and Verhaegen and Verdult (2007). The choice of the particular structure (3) for the input u is limiting in applications in which the input cannot be arbitrarily chosen. We suggest in Section 2.3 a possible way to deal with alternative input signals. Note that the two assumptions imply that $\sigma(A) \cap \sigma(S) = \emptyset$, which in turn implies that Eq. (2) has a unique solution or, equivalently, that the response of system (1) driven by (3) is described by (4). We evaluate Eq. (4) over a set of sample times $T_k^p = \{t_{k-p+1}, \dots, t_{k-1}, t_k\}$ with $0 \leq t_0 < t_1 < \dots < t_{k-p} < \dots < t_k < \dots < t_q$, with $p > 0$ and $q \geq p$. The set T_k^p represents a moving window of p sample times t_i , with $i = 0, \dots, q$. We call Π_k the estimate of the matrix Π at T_k^p , namely the estimate computed at the time t_k using the last p instants of time $t_{k-p+1}, \dots, t_k$. The estimate can be computed as follows.

Theorem 2. Let the time-snapshots $Q_k \in \mathbb{R}^{np \times nv}$ and $\chi_k \in \mathbb{R}^{np}$, with $p \geq v$, be defined as

$$Q_k = \begin{bmatrix} \omega(t_{k-p+1})^\top \otimes I - \omega(0)^\top \otimes e^{At_{k-p+1}} \\ \vdots \\ \omega(t_{k-1})^\top \otimes I - \omega(0)^\top \otimes e^{At_{k-1}} \\ \omega(t_k)^\top \otimes I - \omega(0)^\top \otimes e^{At_k} \end{bmatrix},$$

and

$$\chi_k = \begin{bmatrix} x(t_{k-p+1}) - e^{At_{k-p+1}}x(0) \\ \vdots \\ x(t_{k-1}) - e^{At_{k-1}}x(0) \\ x(t_k) - e^{At_k}x(0) \end{bmatrix},$$

respectively. Assume the matrix Q_k has full column rank, then

$$\text{vec}(\Pi_k) = (Q_k^\top Q_k)^{-1} Q_k^\top \chi_k. \quad (6)$$

Proof. Eq. (4) can be rewritten as

$$\Pi\omega(t) - e^{At}\Pi\omega(0) = x(t) - e^{At}x(0). \quad (7)$$

Using the vectorization operator and the Kronecker product on Eq. (7) yields

$$\text{vec}(\Pi\omega(t)) - \text{vec}(e^{At}\Pi\omega(0)) = \text{vec}(x(t) - e^{At}x(0)),$$

and

$$(\omega(t)^\top \otimes I - \omega(0)^\top \otimes e^{At})\text{vec}(\Pi) = \text{vec}(x(t) - e^{At}x(0)). \quad (8)$$

Computing Eq. (8) at all elements of T_k^p yields

$$Q_k \text{vec}(\Pi_k) = \chi_k. \quad (9)$$

If the matrix Q_k has full column rank, then we can compute Π_k from the last equation yielding Eq. (6). $\square$

Note that the selection of the set T_k^p can affect the *quality* of the data and the rank of the matrix Q_k . Thus, to assure that T_k^p is non-pathological (Chen & Francis, 1995) we introduce the following technical assumption.

Assumption 3. The elements of T_k^v are selected such that $\text{rank}([\omega(t_{k-v+1}) \ \cdots \ \omega(t_k)]) = v$ for all k .

If [Assumption 1](#) holds then it is always possible to choose the elements of T_k^v , i.e. the sampling times, such that [Assumption 3](#) holds (see [Lemma 1](#) in [Padoan, Scarcioiti, & Astolfi, in press](#), for a proof of this fact). We now show that this assumption can be used to make the matrix Q_k full rank.

Lemma 1. Suppose [Assumptions 1–3](#) hold. If $p = v$, then Q_k is square and full rank.

Proof. Let $x(0) \neq \Pi\omega(0)$ and consider the i -block element of the matrix Q_k , namely $\omega(t_i)^\top \otimes I - \omega(0)^\top \otimes e^{At_i}$. Note that the properties of the Kronecker product yield

$$\begin{aligned} \omega(t_i)^\top \otimes I - \omega(0)^\top \otimes e^{At_i} &= \omega(0)^\top e^{S^\top t_i} \otimes I - \omega(0)^\top \otimes I e^{At_i} \\ &= (\omega(0)^\top \otimes I)(e^{S^\top t_i} \otimes I) - (\omega(0)^\top \otimes I)(I \otimes e^{At_i}) \\ &= (\omega(0)^\top \otimes I)(e^{S^\top t_i} \otimes I - I \otimes e^{At_i}). \end{aligned}$$

Since $\sigma(A) \subset \mathbb{C}_{<0}$ and $\sigma(S) \subset \mathbb{C}_0$, the excitability of $(S, \omega(0))$ implies that the i -block element of the matrix Q_k is a $n \times nv$ matrix of rank n . Since the blocks are chosen corresponding to the elements of T_k^v , by [Assumption 3](#) the linear independence holds for all k . As a result Q_k is a square full rank matrix.

Let $x(0) = \Pi\omega(0)$. Since [Eq. \(8\)](#) reduces to $(\omega(t)^\top \otimes I)\text{vec}(\Pi) = \text{vec}(x(t))$, the i -block element of the matrix Q_k is $(\omega(0)^\top \otimes I)(e^{S^\top t_i} \otimes I)$. The claim follows noting that these blocks are chosen according to the elements of T_k^v . Thus by [Assumption 3](#) the blocks are linearly independent for all k . $\square$

Since real data are affected by noise, the assumptions of [Lemma 1](#) may not hold. In this case p , i.e. the number of samples in the moving window, can be selected larger than v and, as well-known from linear algebra and remarked in [Åström and Wittenmark \(1995\)](#) and [Jiang and Jiang \(2012\)](#), the solution of [Eq. \(6\)](#) is the least squares solution of [\(9\)](#). We now prove that the estimate Π_k is actually equal to Π .

Theorem 3. Suppose [Assumptions 1–3](#) hold. Let Π be the solution of [Eq. \(2\)](#). Then Π_k computed by [Eq. \(6\)](#) is equal to Π .

Proof. The matrix Π_k defined in [Eq. \(6\)](#) is such that the equations

$$x(t_k) = \Pi_k \omega(t_k) + e^{At_k}(x(0) - \Pi_k \omega(0)) \quad (10)$$

and

$$\dot{x}(t)|_{t_k} = \Pi_k S \omega(t_k) + A e^{At_k}(x(0) - \Pi_k \omega(0)) \quad (11)$$

hold. Consider the first equation of system [\(1\)](#) computed at t_k , namely

$$\dot{x}(t)|_{t_k} = A x(t_k) + B L \omega(t_k). \quad (12)$$

Substituting [Eqs. \(10\) and \(11\)](#) in [Eq. \(12\)](#) yields

$$\begin{aligned} \Pi_k S \omega(t_k) + A e^{At_k}(x(0) - \Pi_k \omega(0)) \\ = A (\Pi_k \omega(t_k) + e^{At_k}(x(0) - \Pi_k \omega(0))) + B L \omega(t_k), \end{aligned}$$

from which $A \Pi_k + B L = \Pi_k S$ follows. Since $\sigma(A) \cap \sigma(S) = \emptyset$, this equation has a unique solution, hence $\Pi_k = \Pi$. $\square$

The discussion carried out so far has the drawback that information on the state of the system is required. In practice, this is usually not the case and only the output y is available. Thus, we look for a counterpart of [Theorem 2](#) in which the output is used in place of the state.

Theorem 4. Let the time-snapshots $R_k \in \mathbb{R}^{w \times nv}$ and $\gamma_k \in \mathbb{R}^w$, with $w \geq nv$, be defined as

$$R_k = \begin{bmatrix} (\omega(0)^\top \otimes C)(e^{S^\top t_{k-w+1}} \otimes I - I \otimes e^{At_{k-w+1}}) \\ \vdots \\ (\omega(0)^\top \otimes C)(e^{S^\top t_{k-1}} \otimes I - I \otimes e^{At_{k-1}}) \\ (\omega(0)^\top \otimes C)(e^{S^\top t_k} \otimes I - I \otimes e^{At_k}) \end{bmatrix},$$

and

$$\gamma_k = \begin{bmatrix} y(t_{k-w+1}) - C e^{At_{k-w+1}} x(0) \\ \vdots \\ y(t_{k-1}) - C e^{At_{k-1}} x(0) \\ y(t_k) - C e^{At_k} x(0) \end{bmatrix},$$

respectively. Assume the matrix R_k has full column rank, then

$$\text{vec}(\Pi_k) = (R_k^\top R_k)^{-1} R_k^\top \gamma_k. \quad (13)$$

Proof. The result can be proved following the same steps used to obtain [Eq. \(6\)](#). $\square$

Similarly to [Lemma 1](#), the following result guarantees that the matrix R_k is full rank.

Lemma 2. Suppose [Assumptions 1–3](#) hold. If $w = nv$, then R_k is square and full rank.

Proof. The proof is omitted because similar to the one of [Lemma 1](#), although this time also the observability of (C, A) is used. $\square$

2.3. On-line moment estimation from data

[Eq. \(6\)](#) contains terms that depend upon the matrix A and the initial states $x(0)$ and $\omega(0)$. However, exploiting the fact that these terms enter the response as exponentially decaying functions of time, i.e. $\omega(0)^\top \otimes e^{At}$ and $e^{At} x(0)$, we present now an approximate version of the results of the previous section.

Definition 2. Let the time-snapshots $\tilde{Q}_k \in \mathbb{R}^{np \times nv}$ and $\tilde{\chi}_k \in \mathbb{R}^{np}$, with $p \geq v$, be

$$\tilde{Q}_k = [\omega(t_{k-p+1}) \otimes I \ \cdots \ \omega(t_{k-1}) \otimes I \ \omega(t_k) \otimes I]^\top$$

and

$$\tilde{\chi}_k = [x(t_{k-p+1})^\top \ \cdots \ x(t_{k-1})^\top \ x(t_k)^\top]^\top.$$

Assume the matrix $\tilde{Q}_k$ has full column rank, then following the same steps used to obtain [Eq. \(6\)](#), we define

$$\text{vec}(\tilde{\Pi}_k) = (\tilde{Q}_k^\top \tilde{Q}_k)^{-1} \tilde{Q}_k^\top \tilde{\chi}_k. \quad (14)$$

Note that if [Assumption 3](#) holds and $p = v$, then $\tilde{Q}_k$ is square and full rank (the proof of this fact is similar to the proof of [Lemma 1](#) when $x(0) = \Pi\omega(0)$, thus, it is omitted).

We now prove that $\tilde{\Pi}_k$ converges to Π . To this end, we first present a preliminary result.

Lemma 3. Suppose [Assumptions 1–3](#) hold. There exists a matrix $\tilde{\Pi}$ such that $\lim_{t_k \rightarrow \infty} \tilde{\Pi}_k = \tilde{\Pi}$.

Proof. By [Assumptions 1](#) and [2](#) there exists a matrix $\tilde{\Pi}$ such that the steady-state response $x^{ss}(t)$ of the interconnection of system [\(1\)](#) and the generator [\(3\)](#) is described by the equation $x^{ss}(t) = \tilde{\Pi} \omega(t)$. Then substituting $\tilde{\chi}_k^{ss} = \tilde{Q}_k \text{vec}(\tilde{\Pi})$ in [Eq. \(14\)](#), yields

$$\lim_{t_k \rightarrow \infty} \text{vec}(\tilde{\Pi}_k) = (\tilde{Q}_k^\top \tilde{Q}_k)^{-1} \tilde{Q}_k^\top \tilde{\chi}_k^{ss} = \text{vec}(\tilde{\Pi}),$$

which is well-defined by [Assumption 3](#) if $p = v$. $\square$

Theorem 5. Let Π be the solution of Eq. (2). Suppose [Assumptions 1–3](#) hold. There exists a sequence $\{t_k\}$ such that $\lim_{k \rightarrow \infty} \tilde{\Pi}_k = \Pi$.

Proof. The matrix $\tilde{\Pi}_k$ defined in Eq. (14) is such that

$$x(t_k) = \tilde{\Pi}_k \omega(t_k). \quad (15)$$

Substituting Eqs. (15) and (11) in Eq. (12) yields

$$\Pi_k S \omega(t_k) + A e^{At_k} (x(0) - \Pi_k \omega(0)) = A \tilde{\Pi}_k \omega(t_k) + B L \omega(t_k)$$

and since $\Pi_k = \Pi$,

$$(A \tilde{\Pi}_k + B L - \Pi S) \omega(t_k) = A e^{At_k} (x(0) - \Pi \omega(0)),$$

from which, using Eq. (2) and [Assumption 2](#), the equation

$$(\tilde{\Pi}_k - \Pi) \omega(t_k) = e^{At_k} (x(0) - \Pi \omega(0))$$

follows. By [Assumption 1](#) there exists a sequence $\{t_k\}$, with $\lim_{k \rightarrow \infty} t_k = \infty$, such that for any $t_i \in \{t_k\}$, $\omega(t_i) \neq 0$ and [Assumption 3](#) holds. By [Assumption 2](#)

$$\lim_{k \rightarrow \infty} (\tilde{\Pi}_k - \Pi) \omega(t_k) = \lim_{k \rightarrow \infty} e^{At_k} (x(0) - \Pi \omega(0)) = 0,$$

and by [Assumption 3](#) and [Lemma 3](#), $\lim_{k \rightarrow \infty} (\tilde{\Pi}_k - \Pi) = \lim_{k \rightarrow \infty} (\tilde{\Pi} - \Pi) = 0$. It follows that $\tilde{\Pi}_k$ converges asymptotically to Π . $\square$

Remark 1. Eq. (15) is reminiscent of the POD of the collection $\{x(t_i)\}$. However, the two concepts are quite different. In fact, the POD of $\{x(t_i)\}$ is

$$\begin{bmatrix} x(t_0) & \cdots & x(t_q) \end{bmatrix} = \underbrace{\begin{bmatrix} \mu(t_0) & \cdots & \mu(t_q) \end{bmatrix}}_U \Delta$$

with $\Delta \in \mathbb{R}^{q \times q}$ and $U^* U = I$, where the superscript $*$ indicates the complex conjugate transpose. The dimensions of Δ are related to the number of samples, whereas the dimensions of Π are related to the order of the system to be reduced and of the signal generator. In fact, the POD is a decomposition of the entire cloud of data $\{x(t_i)\}$ along the vectors $\mu(t_i)$, called principal directions of $\{x(t_i)\}$ ([Antoulas, 2005](#)). By contrast, in the technique proposed in this paper the oldest data are discarded as soon as new data satisfying [Assumption 3](#) is collected. As a consequence, while Δ is built to describe the entire dynamics of $\{x(t_i)\}$, Π is built to describe only the steady-state response of the system to be reduced. The result is that the POD is usually used with the Petrov–Galerkin projection for a SVD-based approximation ([Rowley et al., 2004](#); [Willcox & Peraire, 2002](#)), whereas this technique is a moment matching method.

A similar discussion can be carried out for Eq. (13) that contains also terms which depend upon the matrix C . In this case note that Eq. (4) can be written as

$$y(t) = C \Pi \omega(t) + \varepsilon(t),$$

with $\varepsilon(t) = C e^{At} (x(0) - \Pi \omega(0))$ an exponentially decaying signal. Thus, an approximate version of [Theorem 4](#) follows.

Theorem 6. Define the time-snapshots $\tilde{R}_k \in \mathbb{R}^{w \times v}$ and $\tilde{\gamma}_k \in \mathbb{R}^w$, with $w \geq v$, as

$$\tilde{R}_k = \begin{bmatrix} \omega(t_{k-w+1}) & \cdots & \omega(t_{k-1}) & \omega(t_k) \end{bmatrix}^\top$$

and

$$\tilde{\gamma}_k = \begin{bmatrix} y(t_{k-w+1}) & \cdots & y(t_{k-1}) & y(t_k) \end{bmatrix}^\top.$$

Assume the matrix $\tilde{R}_k$ has full column rank, then

$$\text{vec}(\tilde{C} \tilde{\Pi}_k) = (\tilde{R}_k^\top \tilde{R}_k)^{-1} \tilde{R}_k^\top \tilde{\gamma}_k, \quad (16)$$

is an approximation of the on-line estimate $C \Pi_k$, namely there exists a sequence $\{t_k\}$ such that $\lim_{k \rightarrow \infty} \tilde{C} \tilde{\Pi}_k = C \Pi$.

Proof. Eq. (16) can be derived following the same steps used to obtain Eq. (6). The convergence of the limit to $C \Pi$ is proved repeating the proof of [Theorem 5](#). $\square$

Note that if [Assumption 3](#) holds and $w = v$, then $\tilde{R}_k$ is square and full rank (the proof of this fact is similar to the proof of [Lemma 1](#), thus, it is omitted). Since $\tilde{R}_k$ is smaller than R_k , the determination of $\tilde{C} \tilde{\Pi}_k$ is computationally less complex than the computation of $\tilde{\Pi}_k$. Note also that, from Eq. (16) we are not able to retrieve the matrix $\tilde{\Pi}_k$, but only $\tilde{C} \tilde{\Pi}_k$. However, as shown in Eq. (5), we only need $C \Pi$ to compute the reduced order model, i.e. Π is not explicitly required. Eq. (16) is a classic least-square estimation formula. Thus, we can provide a recursive formula.

Theorem 7. Assume that $\Phi_k = (\tilde{R}_k^\top \tilde{R}_k)^{-1}$ and $\Psi_k = (\tilde{R}_{k-1}^\top \tilde{R}_{k-1} + \omega(t_k) \omega(t_k)^\top)^{-1}$ are full rank for all $t \geq t_r$ with $t_r \geq t_w$. Given $\text{vec}(\tilde{C} \tilde{\Pi}_r)$, Φ_r and Ψ_r , the recursive least-square formula

$$\begin{aligned} \text{vec}(\tilde{C} \tilde{\Pi}_k) &= \text{vec}(\tilde{C} \tilde{\Pi}_{k-1}) \\ &+ \Phi_k \omega(t_k) (y(t_k) - \omega(t_k)^\top \text{vec}(\tilde{C} \tilde{\Pi}_{k-1})) \\ &- \Phi_k \omega(t_{k-w}) (y(t_{k-w}) - \omega(t_{k-w})^\top \text{vec}(\tilde{C} \tilde{\Pi}_{k-1})), \end{aligned} \quad (17)$$

with

$$\begin{aligned} \Phi_k &= \Psi_k - \Psi_k \omega(t_{k-w}) \\ &\times (I + \omega(t_{k-w})^\top \Psi_k \omega(t_{k-w}))^{-1} \omega(t_{k-w})^\top \Psi_k \end{aligned} \quad (18)$$

and

$$\begin{aligned} \Psi_k &= \Phi_{k-1} - \Phi_{k-1} \omega(t_k) \\ &\times (I + \omega(t_k)^\top \Phi_{k-1} \omega(t_k))^{-1} \omega(t_k)^\top \Phi_{k-1} \end{aligned} \quad (19)$$

holds for all $t \geq t_r$.

Proof. The formula is obtained adapting the results in [Åström and Wittenmark \(1995\)](#) (see also [Ben-Israel & Greville, 2003](#); [Greville, 1960](#); [Wang & Zhang, 2011](#)) to the present scenario, in which at each step we acquire a new measure and we discard an old measure: for completeness we provide the details of the proof. Note that

$$\begin{aligned} \Phi_k^{-1} &= \tilde{R}_k^\top \tilde{R}_k = \sum_{i=k-w+1}^k \omega(t_i) \omega(t_i)^\top \\ &= \sum_{i=k-w}^{k-1} \omega(t_i) \omega(t_i)^\top + \omega(t_k) \omega(t_k)^\top - \omega(t_{k-w}) \omega(t_{k-w})^\top \\ &= \Phi_{k-1}^{-1} + \omega(t_k) \omega(t_k)^\top - \omega(t_{k-w}) \omega(t_{k-w})^\top \end{aligned}$$

and that rewriting Eq. (16) for $k-1$ yields

$$\tilde{R}_{k-1}^\top \tilde{\gamma}_{k-1} = \sum_{i=k-w}^{k-1} \omega(t_i) y(t_i) = \Phi_{k-1}^{-1} \text{vec}(\tilde{C} \tilde{\Pi}_{k-1}).$$

Substituting the first equation in the second we obtain

$$\begin{aligned} \sum_{i=k-w}^{k-1} \omega(t_i) y(t_i) &= \Phi_k^{-1} \text{vec}(\tilde{C} \tilde{\Pi}_{k-1}) \\ &- \omega(t_k) \omega(t_k)^\top \text{vec}(\tilde{C} \tilde{\Pi}_{k-1}) + \omega(t_{k-w}) \omega(t_{k-w})^\top \text{vec}(\tilde{C} \tilde{\Pi}_{k-1}), \end{aligned}$$

which substituted in (16), namely

$$\begin{aligned} \text{vec}(\tilde{C} \tilde{\Pi}_k) &= \Phi_k \left(\sum_{i=k-w}^{k-1} \omega(t_i) y(t_i) + \omega(t_k) y(t_k) - \omega(t_{k-w}) y(t_{k-w}) \right) \end{aligned}$$

yields Eq. (17). Finally Eqs. (18) and (19) are obtained applying recursively the matrix inversion lemma ([Åström & Wittenmark,](#)

1995) to $\Phi_k = (\Psi_k^{-1} - \omega(t_{k-w})\omega(t_{k-w})^\top)^{-1}$ and $\Psi_k = (\Phi_{k-1}^{-1} + \omega(t_k)\omega(t_k)^\top)^{-1}$. $\square$

Note that the construction of the initial values $\text{vec}(\widetilde{C}\widetilde{\Pi}_r)$, Φ_r and Ψ_r needed to start the recursion can be done in two ways: the first consists in using Eq. (16) to build $\text{vec}(\widetilde{C}\widetilde{\Pi}_r)$, Φ_r and Ψ_r and then updating the estimate with the equations in Theorem 7. However, this method has the drawback of requiring the inversion of $(\widetilde{R}_k^\top \widetilde{R}_k)^{-1}$. The second method consists in starting with dummy initial values $\text{vec}(\widetilde{C}\widetilde{\Pi}_r)$, Φ_r and Ψ_r . Since the formulas “forget” the oldest measurements, after a sufficient number of iterations all the dummy measurements are forgotten.

Since for single-input, single-output systems the terms $(I + \omega(t_{k-w})^\top \Psi_k \omega(t_{k-w}))^{-1}$ and $(I + \omega(t_k)^\top \Phi_{k-1} \omega(t_k))^{-1}$ are scalars, updating the matrices Φ_k and Ψ_k does not require any matrix inversion. Hence, Eqs. (17)–(18)–(19) can be used to compute a fast, on-line, estimate of $\widetilde{C}\widetilde{\Pi}_k$, since the computational complexity of updating (17) is $\mathcal{O}(1)$. Thus, the implementation of Eqs. (17)–(18)–(19) to update $\widetilde{C}\widetilde{\Pi}_k$ is preferred to Eq. (16), which has a computational complexity of³ $\mathcal{O}(v^{2.373})$ at each iteration k . In comparison, the Arnoldi or Lanczos procedure for the model reduction by moment matching has a computational complexity of $\mathcal{O}(vn^2)$ (Antoulas, 2005, Section 14.1) (or $\mathcal{O}(\alpha vn)$ for a sparse matrix A , with α the average number of non-zero elements per row/column of A). In addition, note that these procedures require a model to be reduced and thus further expensive computation has to be considered for the identification of the original system.

The approximations $\widetilde{\Pi}_k$ and $\widetilde{C}\widetilde{\Pi}_k$ can be computed with the following algorithm.

Algorithm 1. Let k be a sufficiently large integer. Select $\eta > 0$ sufficiently small. Select $w \geq nv$ ($w \geq v$, respectively).

- 1: Construct the matrices $\widetilde{Q}_k$ and $\widetilde{X}_k$ ($\widetilde{R}_k$ and $\widetilde{T}_k$, respectively).
- 2: **If** $\text{rank}(\widetilde{Q}_k) = nv$ ($\text{rank}(\widetilde{R}_k) = v$, respectively) **then** compute $\widetilde{\Pi}_k$ ($\widetilde{C}\widetilde{\Pi}_k$, respectively) solving Eq. (14) ((16), or (17), respectively).

Else increase w .

If $k - w < 0$ **then** restart the algorithm selecting a larger initial k .

- 3: **If**

$$\|\widetilde{\Pi}_k - \widetilde{\Pi}_{k-1}\| > \frac{\eta}{t_k - t_{k-1}}, \quad (20)$$

$$\left(\|\widetilde{C}\widetilde{\Pi}_k - \widetilde{C}\widetilde{\Pi}_{k-1}\| > \frac{\eta}{t_k - t_{k-1}}, \text{ respectively} \right) \quad (21)$$

then $k = k + 1$ **go to** 1.

- 4: **Stop.**

As already noted, it is more realistic to approximate $\widetilde{C}\widetilde{\Pi}_k$, using output measurements, than to approximate $\widetilde{\Pi}_k$, which needs state measurements. Moreover, the determination of $\widetilde{C}\widetilde{\Pi}_k$ is computationally more efficient since the number of unknown elements is smaller. Nevertheless, the determination of $\widetilde{\Pi}_k$ is not irrelevant. From a theoretical point of view, $\widetilde{\Pi}_k$ provides a way to determine the solution of the Sylvester equation from measurements. Note also that $\widetilde{\Pi}_k$ contains more information than $\widetilde{C}\widetilde{\Pi}_k$. In particular, it provides information regarding the order of the unknown system. Thus, the estimation of $\widetilde{\Pi}_k$ paves the way for an extension of this work in the direction of system identification.

Remark 2. It is not always possible to arbitrarily select the input of the system to be reduced. For instance the input signal may be composed of several frequencies. Instead of system (3), consider the input described by the equations

$$\dot{\omega} = S\omega, \quad u = L\omega + v,$$

with $v(t) \in \mathbb{R}^n$ an unknown signal. In this case the output response of system (1) is

$$y(t) = C\Pi\omega(t) + Ce^{At}(x(0) - \Pi\omega(0)) + \int_0^t e^{A(t-\tau)}Bv(\tau)d\tau,$$

which can be written as

$$y(t) = C\Pi\omega(t) + \varepsilon(t) + v(t),$$

with $v(t) = \int_0^t e^{A(t-\tau)}Bv(\tau)d\tau$ and $\varepsilon(t) = Ce^{At}(x(0) - \Pi\omega(0))$. One can then apply the filtering techniques given in Åström and Wittenmark (1995, Chapter 11): we filter out v from y and u with a band-pass filter and apply the results of the paper to the filtered y_f and u_f .

Remark 3. When the class of inputs is not given, it may be necessary to carry out a preliminary analysis to estimate specific frequencies at which we would like to interpolate the transfer function. This analysis requires an additional step. However, note that the estimation of some of the frequency peaks of the transfer function is not as computationally expensive as a full system identification procedure for two reasons: on one hand we are interested in $v \ll n$ frequency features; on the other hand, since our aim is to obtain a reduced order model, we have the additional benefit of obtaining directly a model of order v (in contrast with the additional computational cost of determining a reduced order model after a high-order identification procedure).

2.4. Families of reduced order models

Using the approximations given by Algorithm 1 a reduced order model of system (1) can be defined at each instant of time t_k .

Definition 3. Consider system (1) and the signal generator (3). Suppose Assumptions 1–3 hold. Then the system

$$\dot{\xi} = F_k\xi + G_ku, \quad \psi = H_k\xi, \quad (22)$$

with $\xi(t) \in \mathbb{R}^v$, $F_k \in \mathbb{R}^{v \times v}$, $G_k \in \mathbb{R}^{v \times 1}$, $H_k \in \mathbb{R}^{1 \times v}$, is a model of system (1) at S at time t_k , if there exists a unique solution P_k of the equation

$$F_kP_k + G_kL = P_kS, \quad (23)$$

such that

$$\widetilde{C}\widetilde{\Pi}_k = H_kP_k, \quad (24)$$

where $\widetilde{C}\widetilde{\Pi}_k$ is the solution of (16).

With this definition we can formulate the following result.

Proposition 1. Select $P_k = I$, for all $k \geq 0$. If $\sigma(F_k) \cap \sigma(S) = \emptyset$ for all $k \geq 0$, then the model

$$\dot{\xi} = (S - G_kL)\xi + G_ku, \quad \psi = \widetilde{C}\widetilde{\Pi}_k\xi, \quad (25)$$

is a model of system (1) at S for all t_k .

Proof. Let $P_k = I$, for all $k \geq 0$. Then Eq. (23) yields $F_k = S - G_kL$ and Eq. (24) yields $H_k = \widetilde{C}\widetilde{\Pi}_k$. $\square$

³ This is the computational complexity of the fastest algorithm (Le Gall, 2014) for the inversion and multiplication of matrices. If the classical Gauss–Jordan elimination is used, then the computational complexity is $\mathcal{O}(v^3)$.

Note that for most purposes, models (22) and (25) can be defined using directly the asymptotic value of $\widetilde{C\Pi}_k$. However, defining the reduced order model at each instant of time t_k allows to implement the algorithm online. This is particularly advantageous when the unknown system has a parameter which is subject to variation. If the variation is sufficiently slow, then the algorithm would be able to produce updated reduced order models at each t_k .

Remark 4. The so-called Loewner framework represents an alternative data-driven approach to model reduction by moment matching (Mayo & Antoulas, 2007). One of the main differences between our approach and the Loewner framework is that we use time-domain measurements, while the Loewner framework makes use of frequency-domain measurements. Thinking in a classical interpolation/Krylov fashion, in the Loewner framework the measurements are used to build projectors with a particular structure. In the framework introduced by Astolfi (2010) the moments can be simply collected in a vector $C\Pi$ which can be used for model reduction directly. Moreover, the use of frequency-domain measurements naturally restricts the applicability of the Loewner framework to linear systems for which frequency measurements are possible. On the other hand, since the proposed technique is based on time-domain measurements, it can be extended to non-linear systems.

Remark 5. The results developed so far can be easily extended to linear time-delay systems. In fact, Algorithm 1 can be used to estimate the moments of linear time-delay systems without any modification. On the other hand the choice of the structure of the time-delay reduced order model is more difficult. We refer the reader to Scarcioiti and Astolfi (2015b, 2016b) for more details.

2.5. Properties of the exponentially converging models

In Astolfi (2010), Ionescu and Astolfi (2013a), Ionescu et al. (2014) and Scarcioiti and Astolfi (2016b) the problem of enforcing additional properties and constraints on the reduced order model has been studied. In this section we go through some of these properties to determine if they hold for the families (25) and under which conditions.

(1) *Matching with prescribed eigenvalues:* Consider system (25) and the problem of determining at every k the matrix G_k such that $\sigma(F_k) = \{\lambda_{1,k}, \dots, \lambda_{v,k}\}$ for some prescribed values $\lambda_{i,k}$. The solution of this problem is well-known and consists in selecting G_k such that $\sigma(S - G_k L) = \sigma(F_k)$. This is possible for every k and for all $\lambda_{i,k} \notin \sigma(S)$ because G_k is independent from the estimate $\widetilde{C\Pi}_k$. Note also that by observability of (L, S) , G_k is unique at every k .

(2) *Matching with prescribed relative degree, matching with prescribed zeros, matching with compartmental constraints:* These problems can be solved at every k as detailed in Astolfi (2010) if and only if

$$\text{rank} \begin{bmatrix} sI - S \\ \widetilde{C\Pi}_k \end{bmatrix} = n, \quad (26)$$

for all $s \in \sigma(S)$ at k . Even though the asymptotic value of $\widetilde{C\Pi}_k$ satisfies this condition there is no guarantee that the condition holds for all k . However, if the condition holds for the asymptotic value, then there exists $\bar{k} \gg 0$ such that for all $k \geq \bar{k}$ condition (26) holds.

2.6. Los Angeles University Hospital building model

In this section we apply Algorithm 1 to the model of a building (Los Angeles University Hospital) with 8 floors, each having

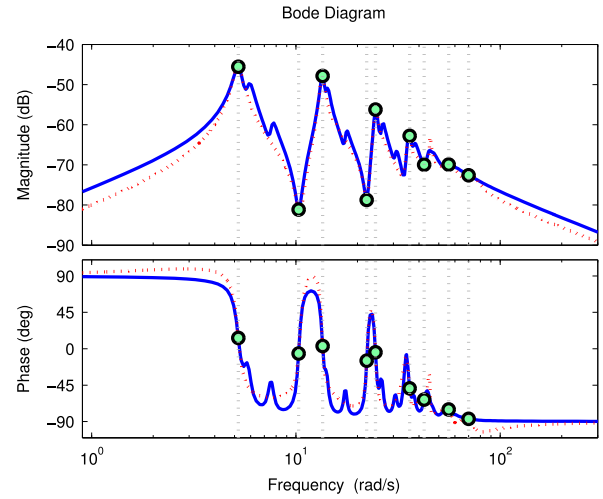


Fig. 1. Bode plot of the building model (solid/blue line) and of the asymptotic reduced-order model (dotted/red line). The circles indicate the interpolation points.

three degrees of freedom (Antoulas, Sorensen, & Gugercin, 2001; Chahlaoui & Van Dooren, 2005). The model is described by equations of the form (1) with a state of dimension $n = 48$. The output of the system is the motion of the first coordinate. Note that this model has been reduced with various methods in Antoulas (2005), obtaining a reduced-order model of order $v = 31$. In this paper we reduce the system with a model of order $v = 19$, interpolating at the points $0, \pm 5.22i, \pm 10.3i, \pm 13.5i, \pm 22.2i, \pm 24.5i, \pm 36i, \pm 42.4i, \pm 55.9i$ and $\pm 70i$, corresponding to the main peaks in the magnitude of the frequency response of the system. A reduced order model (25) at time t_k has been constructed assigning the eigenvalues of F_k (determined with the data-driven technique given in Scarcioiti, Jiang, & Astolfi, 2016). Fig. 1 shows the Bode plot of the system (solid/blue line) and of the estimated reduced-order model at $t = 100$ (dotted/red line). The green circles indicate the interpolation points. The surface in Fig. 2 (Fig. 3, respectively) represents the magnitude (phase, respectively) of the transfer function of the reduced order model as a function of t_k , with $4.1 \leq t_k \leq 23.7$ s. The solid/blue line indicates the magnitude (phase, respectively) of the transfer function of the reduced order model for the exact moments $C\Pi$. The figures show how the approximated magnitude and phase of model (25) at S of system (1) evolve over time and approach the exact reduced order model as $t_k \rightarrow \infty$. We also check the qualitative behavior of the reduced order model when the input is not an interpolating signal. To this end, we select the input

$$u = c_1 \cos(0.001t) + c_2 \sin(0.1t) + c_3 \sin(3t) + c_4 \cos(15t) + c_5 \sin(23t) + c_6 \sin(37t) + c_7 \sin(50t), \quad (27)$$

where c_i , with $i = 1, \dots, 7$, are randomly generated weights such that $\sum_{i=1}^7 c_i = 1$. Fig. 4 (top graph) shows the output response of the building model in solid/blue line and the output response of the asymptotic reduced order model in dotted/red line when the input to the two systems is selected as in (27). The bottom graph shows the normalized absolute error between the two output responses, i.e. $|y(t) - \Psi_\infty(t)| / \max(y(t))$. We see that the error is fairly small even though the input is not an interpolating signal.

3. Nonlinear systems

3.1. Model reduction by moment matching – recalled

Similarly to the linear case, to render the paper self-contained we recall the notion of moment for nonlinear systems as presented

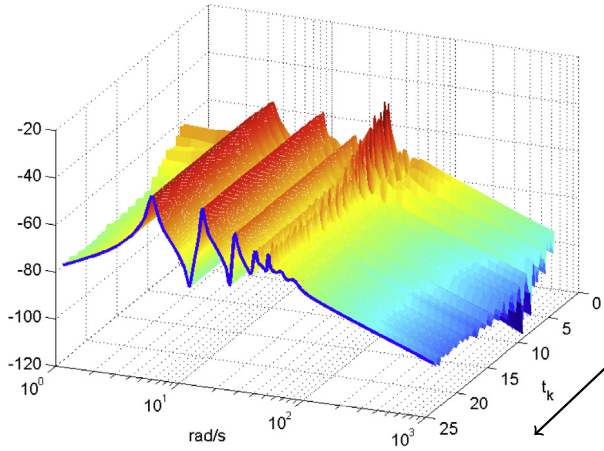


Fig. 2. The mesh represents the magnitude of the transfer function of the reduced order model as a function of t_k , with $4.1 \leq t_k \leq 23.7$ s. The solid/blue line indicates the magnitude of the transfer function of the reduced order model for the moments $C\Pi$.

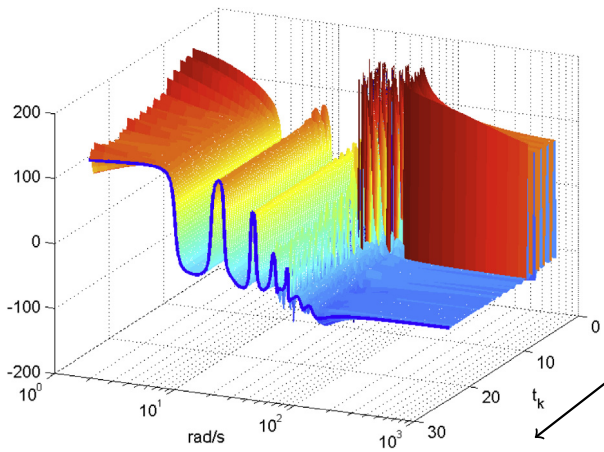


Fig. 3. The mesh represents the phase of the transfer function of the reduced order model as a function of t_k , with $4.1 \leq t_k \leq 23.7$ s. The solid/blue line indicates the phase of the transfer function of the reduced order model for the moments $C\Pi$.

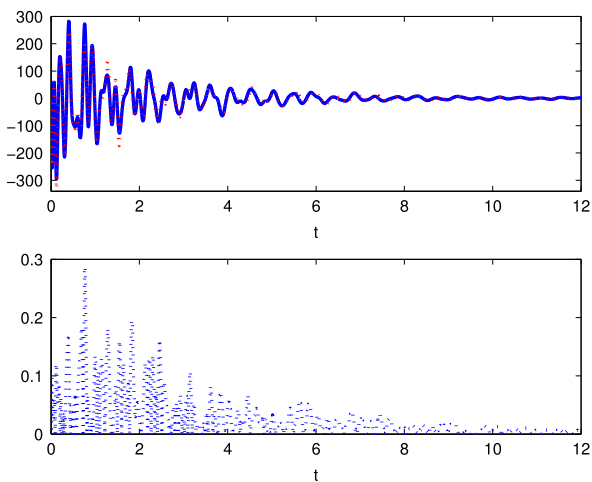


Fig. 4. Top graph: time histories of the output response of the building model (solid/blue line) and of the output response of the asymptotic reduced-order model (dotted/red line) with $u \neq L\omega$. Bottom graph: normalized absolute error between the two outputs.

in Astolfi (2010). Consider a nonlinear, single-input, single-output, continuous-time, system described by the equations

$$\dot{x} = f(x, u), \quad y = h(x), \quad (28)$$

with $x(t) \in \mathbb{R}^n$, $u(t) \in \mathbb{R}$, $y(t) \in \mathbb{R}$, f and h smooth mappings, a signal generator described by the equations

$$\dot{\omega} = s(\omega), \quad u = l(\omega), \quad (29)$$

with $\omega(t) \in \mathbb{R}^v$, s and l smooth mappings, and the interconnected system

$$\dot{\omega} = s(\omega), \quad \dot{x} = f(x, l(\omega)), \quad y = h(x). \quad (30)$$

In addition, suppose that $f(0, 0) = 0$, $s(0) = 0$, $l(0) = 0$ and $h(0) = 0$.

The following two assumptions play, in the nonlinear framework, the role that Assumptions 1 and 2 play in the linear case.

Assumption 4. The signal generator (29) is observable⁴ and neutrally stable.⁵

Assumption 5. The zero equilibrium of the system $\dot{x} = f(x, 0)$ is locally exponentially stable.

These two assumptions can be used to give a description of steady-state response for the nonlinear system (28).

Lemma 4 (Isidori, 1995). Consider system (28) and the signal generator (29). Suppose Assumptions 4 and 5 hold. Then there is a mapping π , locally defined in a neighborhood of $\omega = 0$, which solves the partial differential equation

$$\frac{\partial \pi}{\partial \omega} s(\omega) = f(\pi(\omega), l(\omega)). \quad (31)$$

In addition, for any $x(0)$ and $\omega(0)$ sufficiently small, the steady-state response of system (30) is $\pi(\omega)$.

Lemma 4 implies that the interconnected system (30) possesses an invariant manifold described by the equation $x = \pi(\omega)$, which can be used to define the moment for nonlinear systems.

Definition 4. Consider system (28) and the signal generator (29). Suppose Assumption 4 holds. The function $h \circ \pi$, with π solution of Eq. (31), is the *moment of system (28) at (s, l)* .

Note that the moment $h \circ \pi$ computed along a particular trajectory $\omega(t)$ coincides with the steady-state response of system (28) driven by (29).

Finally, as shown in Astolfi (2010), the system described by the equations

$$\dot{\xi} = s(\xi) - \delta(\xi)l(\xi) + \delta(\xi)u, \quad \psi = h(\pi(\xi)), \quad (32)$$

where δ is any mapping such that the equation

$$\frac{\partial p}{\partial \omega} s(\omega) = s(p(\omega)) - \delta(p(\omega))l(p(\omega)) + \delta(p(\omega))l(\omega),$$

has the unique solution $p(\omega) = \omega$, is a family of *reduced order models of (28) at (s, l)* .

⁴ System (29) is observable if for any pair of initial conditions $\omega_a(0)$ and $\omega_b(0)$, such that $\omega_a(0) \neq \omega_b(0)$, the corresponding output trajectories $l(\omega_a(t))$ and $l(\omega_b(t))$ are such that $l(\omega_a(t)) - l(\omega_b(t)) \neq 0$.

⁵ See Isidori (1995, Chapter 8) for the definition of neutral stability.

3.2. On-line moment estimation from data

Solving Eq. (31) with respect to the mapping π is a difficult task even when there is perfect knowledge of the dynamics of the system, i.e. the mapping f . When f is not known Eq. (31) can be solved numerically, however this has the additional drawback of requiring information on the state of the system. In practice, this is usually not the case and only the output y is available, with the consequence that also the mapping h has to be known.

Note that given the exponential stability hypothesis on the system and Lemma 4, the equation

$$y(t) = h(\pi(\omega(t))) + \varepsilon(t), \quad (33)$$

where $\varepsilon(t)$ is an exponentially decaying signal, holds. We introduce the following assumption.

Assumption 6. The mapping $h \circ \pi$ belongs to the function space identified by the family of continuous basis functions $\varphi_j : \mathbb{R}^v \rightarrow \mathbb{R}$, with $j = 1, \dots, M$ (M may be ∞), i.e. there exist constants $\gamma_j \in \mathbb{R}$, with $j = 1, \dots, M$, such that $h(\pi(\omega)) = \sum_{j=1}^M \gamma_j \varphi_j(\omega)$, for any ω .

The assumption that the mapping to be approximated can be represented by a family of basis functions is standard, see e.g. Toth (2010). For some families of basis functions, e.g. radial basis functions, there exist results of “universal” approximation (Park & Sandberg, 1991; Rocha, 2009). Practically a trial and error procedure can be implemented, for instance starting with the use of a polynomial expansion or with the use of an expansion based on functions belonging to the same class as the ones generated by the signal generator (e.g. sinusoids, for sinusoidal inputs).

Thus, let

$$\Gamma = [\gamma_1 \ \gamma_2 \ \dots \ \gamma_N],$$

$$\Omega(\omega(t)) = [\varphi_1(\omega(t)) \ \varphi_2(\omega(t)) \ \dots \ \varphi_N(\omega(t))]^\top,$$

with $N \leq M$. Using a weighted sum of basis functions, Eq. (33) can be written as

$$y(t) = \sum_{j=1}^N \gamma_j \varphi_j(\omega(t)) + e(t) + \varepsilon(t)$$

$$= \Gamma \Omega(\omega(t)) + e(t) + \varepsilon(t), \quad (34)$$

where $e(t) = \sum_{j=N+1}^M \gamma_j \varphi_j(\omega(t))$ is the error resulting by terminating the summation at N . Consider now the approximation

$$y(t) \approx \sum_{j=1}^N \tilde{\gamma}_j \varphi_j(\omega(t)) = \tilde{\Gamma} \Omega(\omega(t)), \quad (35)$$

which neglects the error $e(t)$ and the transient error $\varepsilon(t)$. Let $\tilde{\Gamma}_k$ be an on-line estimate of the matrix Γ computed at T_k^w , namely computed at the time t_k using the last w instants of time t_i assuming that $e(t)$ and $\varepsilon(t)$ are known. Since this is not the case in practice, define $\tilde{\Gamma}_k$ as the approximation, in the sense of (35), of the estimate $\tilde{\Gamma}_k$. Finally we can compute this approximation as follows.

Theorem 8. Define the time-snapshots $\tilde{U}_k \in \mathbb{R}^{w \times N}$ and $\tilde{\gamma}_k \in \mathbb{R}^w$, with $w \geq v$, as

$$\tilde{U}_k = [\Omega(\omega(t_{k-w+1})) \ \dots \ \Omega(\omega(t_{k-1})) \ \Omega(\omega(t_k))]^\top \quad (36)$$

and

$$\tilde{\gamma}_k = [y(t_{k-w+1}) \ \dots \ y(t_{k-1}) \ y(t_k)]^\top. \quad (37)$$

If $\tilde{U}_k$ is full column rank, then

$$\text{vec}(\tilde{\Gamma}_k) = (\tilde{U}_k^\top \tilde{U}_k)^{-1} \tilde{U}_k^\top \tilde{\gamma}_k, \quad (38)$$

is an approximation of the estimate Γ_k .

To ensure that the approximation is well-defined for all k , we need that the elements of T_k^w be selected such that $\tilde{U}_k^\top \tilde{U}_k$ is full column rank. This condition expresses a property of persistence of excitation that is guaranteed by the following assumption (Padoan et al., 2016).

Assumption 7. The initial condition $\omega(0)$ of system (29) is almost periodic⁶ and all the solutions of the system are analytic. In addition, system (29) satisfies the excitation rank condition⁷ at $\omega(0)$.

To ease the notation we introduce the following definition.

Definition 5. The estimated moment of system (28) is defined as

$$\widetilde{h \circ \pi}_{N,k}(\omega(t)) = \tilde{\Gamma}_k \Omega(\omega(t)), \quad (39)$$

for all $t \in \mathbb{R}$, with $\tilde{\Gamma}_k$ computed using (38).

Note that a nonlinear counterpart of Theorem 7 can also be formulated. The recursive least-square algorithm is obtained with $\Omega(\omega(t_k))$ playing the role of $\omega(t_k)$, $\tilde{U}_k$ playing the role of $\tilde{R}_k$ and $\tilde{\Gamma}_k$ playing the role of $\tilde{C}\tilde{\Pi}_k$, namely

$$\begin{aligned} \text{vec}(\tilde{\Gamma}_k) &= \text{vec}(\tilde{\Gamma}_{k-1}) \\ &+ \Phi_k \Omega(\omega(t_k)) (y(t_k) - \Omega(\omega(t_k))^\top \text{vec}(\tilde{\Gamma}_{k-1})) \\ &- \Phi_k \Omega(\omega(t_{k-w})) (y(t_{k-w}) - \Omega(\omega(t_{k-w}))^\top \text{vec}(\tilde{\Gamma}_{k-1})), \end{aligned} \quad (40)$$

with $\Phi_k = (\tilde{U}_k^\top \tilde{U}_k)^{-1}$ and $\Psi_k = (\tilde{U}_{k-1}^\top \tilde{U}_{k-1} + \Omega(\omega(t_k)) \Omega(\omega(t_k))^\top)^{-1}$. Similarly, Algorithm 1 can be adapted to the present scenario to determine the approximation $\widetilde{h \circ \pi}_{N,k}$ (if the system is linear, Algorithm 1 is recovered selecting $N = v$ and $\phi_j = \varepsilon_j$).

Algorithm 2. Let k be a sufficiently large integer. Select $\eta > 0$ sufficiently small. Select $w \geq v$.

- 1: Construct the matrices $\tilde{U}_k$ and $\tilde{\gamma}_k$.
- 2: If $\text{rank}(\tilde{U}_k) = N$ then compute $\tilde{\Gamma}_k$ solving Eq. (38) (or (40)).
Else increase w .
If $k - w < 0$ then restart the algorithm selecting a larger initial k .
- 3: If $\|\tilde{\Gamma}_k - \tilde{\Gamma}_{k-1}\| > \frac{\eta}{t_k - t_{k-1}}$,
then $k = k + 1$ go to 1.
- 4: Stop.

The convergence of the estimated moment is guaranteed by the next result.

Theorem 9. Suppose Assumptions 4–7 hold. Then $\lim_{t \rightarrow \infty} (h(\pi(\omega(t))) - \lim_{N \rightarrow M} \widetilde{h \circ \pi}_{N,k}(\omega(t))) = 0$.

Proof. Assumption 7 guarantees that the approximation $\tilde{\Gamma}_k$ is well-defined for all k , whereas Assumptions 4 and 5 guarantee that Lemma 4 holds and thus that $h \circ \pi$ is well-defined. The quantity $\|\varepsilon(t_k)\|$ vanishes exponentially by Assumption 5. Hence, by Assumption 6, $\lim_{N \rightarrow M} \widetilde{h \circ \pi}_{N,k}(\omega(t))$ converges to $h(\pi(\omega(t)))$. $\square$

Note that if we try to apply the linear algorithm to data generated by a nonlinear system, then the algorithm would not converge. In

⁶ See Padoan et al. (2016) for the definition of almost periodic point.

⁷ See Padoan et al. (2016) for the definition of excitation rank condition.

fact, the algorithm would try to encode in a linear term information regarding higher order terms. Hence, conditions (20) or (21) would not be satisfied.

Up to this point we have always considered one trajectory $\omega(t)$. While this is sufficient in a linear setting, in which local properties are also global, it may be restrictive in the nonlinear setting. To solve this issue, Algorithm 2 can be easily modified to operate with multiple trajectories. To this end, it suffices to implement the algorithm replacing the matrices $\tilde{U}_k$ and $\tilde{\gamma}_k$ with the matrices

$$\mathcal{U} = [\tilde{U}_k^{1\top} \quad \tilde{U}_k^{2\top} \quad \dots \quad \tilde{U}_k^{q\top}]^\top, \quad (42)$$

$$\mathcal{Y} = [\tilde{\gamma}_k^{1\top} \quad \tilde{\gamma}_k^{2\top} \quad \dots \quad \tilde{\gamma}_k^{q\top}]^\top,$$

respectively, where $\tilde{U}_k^i$ and $\tilde{\gamma}_k^i$ are the matrices in (36) and (37), respectively, sampled along the trajectory of system (29) starting from the initial condition $\omega(0) = \omega_0^i$, with $q \geq 1$. We refer to this method as the “ $\mathcal{U}/\mathcal{Y}$ ” variation.

3.3. Families of reduced order models

Using the approximation given by (39) a reduced order model of system (28) can be defined at each instant of time t_k .

Definition 6. Consider system (28) and the signal generator (29). Suppose Assumptions 4–7 hold. Then the system

$$\dot{\xi} = \phi_k(\xi, u), \quad \psi = \kappa_k(\xi), \quad (43)$$

with $\xi(t) \in \mathbb{R}^v$, $u(t) \in \mathbb{R}$, $\psi(t) \in \mathbb{R}$ and ϕ_k and κ_k smooth mappings, is a model of system (28) at (s, l) at time t_k , i.e. system (43) has the same moment of system (28) at (s, l) , if the equation

$$\frac{\partial p_k}{\partial \omega} s(\omega) = \phi_k(p_k(\omega), l(\omega)), \quad (44)$$

has a unique solution p_k such that

$$\widetilde{h \circ \pi_{N,k}(\omega)} = \kappa_k(p_k(\omega)), \quad (45)$$

where $\widetilde{h \circ \pi_{N,k}(\omega)}$ is obtained by (38).

The next result is a direct consequence of Definition 6 and Astolfi (2010).

Proposition 2. Let δ_k be such that, for all $k \geq 0$,

$$\frac{\partial p_k}{\partial \omega} s(\omega) = s(p_k(\omega)) - \delta_k(p_k(\omega))l(p_k(\omega)) + \delta_k(p_k(\omega))l(\omega),$$

has the unique solution $p_k(\omega) = \omega$ and select $\kappa_k(\omega) = \widetilde{h \circ \pi_{N,k}(\omega)}$. Then the system described by the equations

$$\begin{aligned} \dot{\xi} &= \widetilde{s(\xi)} - \delta_k(\xi)l(\xi) + \delta_k(\xi)u, \\ \psi &= \widetilde{h \circ \pi_{N,k}(\xi)}, \end{aligned} \quad (46)$$

is a model of system (28) at (s, l) for all t_k .

Similarly to the linear case (see Section 2.5), the conditions given in Astolfi (2010) to enforce additional properties upon the reduced order model can be adapted to hold in the present scenario. Moreover, the results of this section can be extended to time-delay systems. In fact, Algorithm 2 can be used to estimate the moments of nonlinear time-delay systems without any modification, see Scarcioiti and Astolfi (2015c, 2016b).

3.4. An approximated nonlinear reduced order model of the DC-to-DC Ćuk converter

In this section we revisit the example given in Scarcioiti and Astolfi (2015c). Therein, using a linear signal generator, a reduced

order model of the DC-to-DC Ćuk converter with linear state dynamics has been given. Herein, we obtain a planar reduced order model with nonlinear state dynamics using an input generated by a nonlinear mapping l . Note that since the system is of low-order, the example serves only illustrative purposes. In particular, we provide a proof of principle that the method allows to obtain a nonlinear reduced order model without solving a partial differential equation.

The averaged model of the DC-to-DC Ćuk converter is given by the equations (Rodriguez, Ortega, & Astolfi, 2005)

$$\begin{aligned} L_1 \dot{x}_1 &= -(1-u)x_2 + E, & C_2 \dot{x}_2 &= (1-u)x_1 + ux_3, \\ L_3 \dot{x}_3 &= -ux_2 - x_4, & C_4 \dot{x}_4 &= x_3 - Gx_4, \\ y &= x_4, \end{aligned} \quad (47)$$

where $x_1(t) \in \mathbb{R}_{\geq 0}$ and $x_3(t) \in \mathbb{R}_{\leq 0}$ describe the averaged currents, $x_2(t) \in \mathbb{R}_{\geq 0}$ and $x_4(t) \in \mathbb{R}_{\leq 0}$ the averaged voltages, L_1, C_2, L_3, E and G positive parameters and $u(t) \in (0, 1)$ a continuous control signal which represents the slew rate of a pulse width modulation circuit used to control the switch position in the converter. In the remaining of the paper we used the numerical values given in Rodriguez et al. (2005) to simulate system (47). Consider the input generated by the equations

$$\begin{aligned} \dot{\omega}_1 &= -75\omega_2, & \dot{\omega}_2 &= 75\omega_1, \\ u &= \max\left(0.15, \frac{1}{2}\left(\omega_1\omega_2 + \omega_1^3 + \frac{1}{2}\right)\right), \end{aligned}$$

which generates a positive input signal with higher order harmonics.

Algorithm 2 has been applied with $t_j = j \cdot 10^{-4}$, $w = 838$ and $\eta = 1.2 \cdot 10^{-7}$. The basis functions are selected as the polynomial surfaces of degree 2 in the first argument and 5 in the second, namely $N = 15$, $\varphi_1(\omega) = 1$, $\varphi_2(\omega) = \omega_1$, $\varphi_3(\omega) = \omega_2$, $\dots$, $\varphi_{15}(\omega) = \omega_2^5$. To improve the quality of the approximation the “ $\mathcal{U}/\mathcal{Y}$ ” variation of the algorithm has been implemented selecting $\omega_0^i \in W := [-0.6, 0.6] \times [-0.6, 0.6]$. These values of ω_0^i generate inputs u in the set⁸ $[0.15, 0.62]$. The final value of k is 4000, for which the algorithm stops with $\eta = 1.186 \cdot 10^{-7}$. The resulting estimated moment is

$$\begin{aligned} \widetilde{h \circ \pi(\omega)}_{15,\eta} &= -3.781 - 3.116\omega_1 + 0.641\omega_2 - 1.71\omega_1^2 \\ &\quad - 9.964\omega_1\omega_2 - 1.243\omega_2^2 - 12.23\omega_1^2\omega_2 + 5.185\omega_1\omega_2^2 \\ &\quad - 1.764\omega_2^3 - 6.876\omega_1^2\omega_2^2 - 3.026\omega_1\omega_2^3 + 0.8862\omega_2^4 \\ &\quad + 14.63\omega_1^2\omega_2^3 - 1.709\omega_1\omega_2^4 + 1.173\omega_2^5. \end{aligned} \quad (48)$$

This polynomial approximation fits well for $u < 0.8$, whereas it does not decrease as fast as the actual output of the system when the input is close to 1: for this value the output of the converter becomes negatively unbounded. This suggests that the following results can be improved if other basis functions are used.

The reduced order model is chosen as in Proposition 2, with $\delta_\eta = 220 \begin{bmatrix} 1 & 1 \end{bmatrix}^\top$, namely

$$\begin{aligned} \dot{\xi}_1 &= 75\xi_2 + 220 \left(u - \max\left(0.15, \frac{1}{2}\left(\xi_1\xi_2 + \xi_1^3 + \frac{1}{2}\right)\right) \right), \\ \dot{\xi}_2 &= -75\xi_1 + 220 \left(u - \max\left(0.15, \frac{1}{2}\left(\xi_1\xi_2 + \xi_1^3 + \frac{1}{2}\right)\right) \right), \\ \psi &= \widetilde{h \circ \pi(\xi)}_{15,\eta}. \end{aligned}$$

The top graph in Fig. 5 shows the time histories of the output of system (47) (solid/blue line) and of the reduced order model (dotted/red line) for the input sequence represented in the bottom

⁸ Simulations showed that polynomial surfaces of higher order are necessary to have a satisfactory approximation in a larger region W .

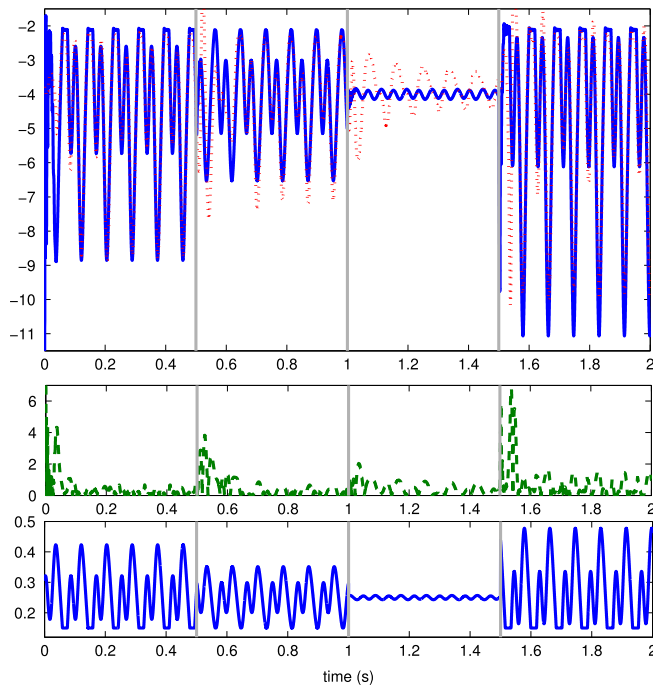


Fig. 5. Top graph: time histories of the output of system (47) (solid/blue line) and of the nonlinear reduced order model (dotted/red line) for the input sequence represented in the bottom graph. The switching times are indicated by solid/gray vertical lines. Middle graph: absolute error (dashed/green line) between the two outputs.

graph. The input is obtained switching $\omega(0)$ every 0.5s (the switching times are indicated by solid/gray vertical lines). $\omega(0)$ takes, in order, the values of $(-0.45, -0.45)$, $(-0.25, -0.45)$, $(0.15, 0.05)$ and $(0.5, 0.5)$. The middle graph in Fig. 5 shows the absolute error (dashed/green line) between the two outputs. We note that the error is mostly due to the poor transient performance. This problem could be alleviated with a selection of δ_η as a function of ξ . The already small steady-state error could be further reduced with a different selection of basis functions.

4. Conclusion

We have presented a theoretical framework and a collection of techniques to obtain reduced order models by moment matching from input/output data for linear systems and nonlinear systems. The approximations proposed in the paper have been exploited to construct families of reduced order models. We have shown that these models asymptotically match the moments of the unknown system to be reduced. Conditions to enforce additional properties upon the reduced order model have been discussed. The use of the algorithm is illustrated by several examples.

References

- Adamjan, V. M., Arov, D. Z., & Krein, M. G. (1971). Analytic properties of Schmidt pairs for a Hankel operator and the generalized Schur–Takagi problem. *Mathematics of the USSR Sbornik*, 15, 31–73.
- Al-Baiyat, S. A., Bettayeb, M., & Al-Saggaf, U. M. (1994). New model reduction scheme for bilinear systems. *International Journal of Systems Science*, 25(10), 1631–1642.
- Antoulas, A. (2005). *Advances in design and control, Approximation of large-scale dynamical systems*. Philadelphia, PA: SIAM.
- Antoulas, A. C., Ball, J. A., Kang, J., & Willems, J. C. (1990). On the solution of the minimal rational interpolation problem. *Linear Algebra and Its Applications*, 137–138, 511–573. Special issue on matrix problems.
- Antoulas, A. C., Sorensen, D. C., & Gugercin, S. (2001). A survey of model reduction methods for large-scale systems. *Contemporary Mathematics*, 280, 193–219.
- Astolfi, A. (2007). A new look at model reduction by moment matching for linear systems. In *Proceedings of the 46th IEEE conference on decision and control*, Dec. (pp. 4361–4366).

- Astolfi, A. (2010). Model reduction by moment matching for linear and nonlinear systems. *IEEE Transactions on Automatic Control*, 55(10), 2321–2336.
- Astrid, P., Weiland, S., Willcox, K., & Backx, T. (2008). Missing point estimation in models described by proper orthogonal decomposition. *IEEE Transactions on Automatic Control*, 53(10), 2237–2251.
- Åström, K. J., & Kumar, P. R. (2014). Control: A perspective. *Automatica*, 50(1), 3–43.
- Åström, K. J., & Wittenmark, B. (1995). *Addison-Wesley series in electrical engineering: control engineering, Adaptive control*.
- Baird, L. C. (1994). Reinforcement learning in continuous time: advantage updating. *Jun. 4* (pp. 2448–2453).
- Beattie, C. A., & Gugercin, S. (2008). Interpolation theory for structure-preserving model reduction. In *Proceedings of the 47th IEEE conference on decision and control*, Cancun, Mexico.
- Ben-Israel, A., & Greville, T. N. E. (2003). *CMS books in mathematics, Generalized inverses: Theory and applications*. Springer.
- Besselink, B., van de Wouw, N., & Nijmeijer, H. (2012). Model reduction for a class of convergent nonlinear systems. *IEEE Transactions on Automatic Control*, 57(4), 1071–1076.
- Besselink, B., van de Wouw, N., & Nijmeijer, H. (2013). Model reduction for nonlinear systems with incremental gain or passivity properties. *Automatica*, 49(4), 861–872.
- Besselink, B., van de Wouw, N., Scherpen, J. M. A., & Nijmeijer, H. (2014). Model reduction for nonlinear systems by incremental balanced truncation. *IEEE Transactions on Automatic Control*, 59(10), 2739–2753.
- Bian, T., Jiang, Y., & Jiang, Z. P. (2014). Adaptive dynamic programming and optimal control of nonlinear nonaffine systems. *Automatica*, 50(10), 2624–2632.
- Byrnes, C. I., Lindquist, A., & Georgiou, T. T. (2001). A generalized entropy criterion for Nevanlinna–Pick interpolation with degree constraint. *IEEE Transactions on Automatic Control*, 46, 822–839.
- Byrnes, C. I., Lindquist, A., Gusev, S. V., & Matveev, A. S. (1995). A complete parameterization of all positive rational extensions of a covariance sequence. *IEEE Transactions on Automatic Control*, 40, 1841–1857.
- Chahlaoui, Y., & Van Dooren, P. (2005). *Dimension reduction of large-scale systems: Proceedings of a workshop*. Berlin, Heidelberg: Springer, Chapter Benchmark examples for model reduction of linear time-invariant dynamical systems.
- Chen, T., & Francis, B. A. (1995). *Optimal sampled-data control systems*. Springer.
- Cooper, J. E. (1999). On-line version of the eigensystem realization algorithm using data correlations. *Journal of Guidance, Control, and Dynamics*, 20(1), 137–142.
- Fujimoto, K. (2008). Balanced realization and model order reduction for port-Hamiltonian systems. *Journal of System Design and Dynamics*, 2(3), 694–702.
- Fujimoto, K., & Scherpen, J. M. A. (2010). Balanced realization and model order reduction for nonlinear systems based on singular value analysis. *SIAM Journal on Control and Optimization*, 48(7), 4591–4623.
- Fujimoto, K., & Tsubakino, D. (2008). Computation of nonlinear balanced realization and model reduction based on Taylor series expansion. *Systems & Control Letters*, 57(4), 283–289.
- Gallivan, K., Vandendorpe, A., & Van Dooren, P. (2004). Sylvester equations and projection-based model reduction. *Journal of Computational and Applied Mathematics*, 162(1), 213–229.
- Gallivan, K. A., Vandendorpe, A., & Van Dooren, P. (2006). Model reduction and the solution of Sylvester equations. In *MTNS, Kyoto*.
- Georgiou, T. T. (1999). The interpolation problem with a degree constraint. *IEEE Transactions on Automatic Control*, 44, 631–635.
- Glover, K. (1984). All optimal Hankel-norm approximations of linear multivariable systems and their L^∞ -error bounds. *International Journal of Control*, 39, 1115–1193.
- Gray, W. S., & Mesko, J. (1997). General input balancing and model reduction for linear and nonlinear systems. In *European control conference*, Brussels, Belgium.
- Gray, W. S., & Mesko, J. (1999). Observability functions for linear and nonlinear systems. *Systems & Control Letters*, 38(2), 99–113.
- Gray, W. S., & Scherpen, J. M. A. (2005). Hankel singular value functions from Schmidt pairs for nonlinear input–output systems. *Systems & Control Letters*, 54(2), 135–144.
- Gray, W. S., & Verriest, E. I. (2006). Balanced realizations near stable invariant manifolds. *Automatica*, 42(4), 653–659.
- Greville, T. N. E. (1960). Some applications of the pseudoinverse of a matrix. *SIAM Review*, 2, 15–22.
- Gugercin, S., Antoulas, A. C., & Beattie, C. (2008). $\mathcal{H}_2$ model reduction for large-scale linear dynamical systems. *SIAM Journal on Matrix Analysis and Applications*, 30(2), 609–638.
- Hemati, M. S., Williams, M. O., & Rowley, C. W. (2014). Dynamic mode decomposition for large and streaming datasets. *Physics of Fluids*, 26(11701), 1–6.
- Hinze, M., & Volkwein, S. (2005). Proper orthogonal decomposition surrogate models for nonlinear dynamical systems: Error estimates and suboptimal control. In *Lecture notes in computational and applied mathematics, Dimension reduction of large-scale systems* (pp. 261–306). Springer.
- Houtzager, I., van Wingerden, J. W., & Verhaegen, M. (2012). Recursive predictor-based subspace identification with application to the real-time closed-loop tracking of flutter. *IEEE Transactions on Control Systems Technology*, 20(4), 934–949.
- Ionescu, T. C., & Astolfi, A. (2013a). Families of moment matching based, structure preserving approximations for linear port Hamiltonian systems. *Automatica*, 49, 2424–2434.
- Ionescu, T. C., & Astolfi, A. (2013b). Families of reduced order models that achieve nonlinear moment matching. In *Proceedings of the 2013 American control conference*, Washington, DC, USA, June 17–19 (pp. 5518–5523).

- Ionescu, T. C., Astolfi, A., & Colaneri, P. (2014). Families of moment matching based, low order approximations for linear systems. *Systems & Control Letters*, 64, 47–56.
- Isidori, A. (1995). *Communications and control engineering, Nonlinear control systems* (3rd Ed.). Springer.
- Jiang, Y., & Jiang, Z. P. (2012). Computational adaptive optimal control for continuous-time linear systems with completely unknown dynamics. *Automatica*, 48(10), 2699–2704.
- Jiang, Y., & Jiang, Z. P. (2014). Robust adaptive dynamic programming and feedback stabilization of nonlinear systems. *IEEE Transactions on Neural Networks and Learning Systems*, 25(5), 882–893.
- Kimura, H. (1986). Positive partial realization of covariance sequences. *Modeling, identification and robust control*, 499–513.
- Kunisch, K., & Volkwein, S. (1999). Control of the Burgers equation by a reduced-order approach using proper orthogonal decomposition. *Journal of Optimization Theory and Applications*, 102(2), 345–371.
- Kunisch, K., & Volkwein, S. (2008). Proper orthogonal decomposition for optimality systems. *ESAIM: Mathematical Modelling and Numerical Analysis*, 42(01), 1–23.
- Lall, S., & Beck, C. (2003). Error bounds for balanced model reduction of linear time-varying systems. *IEEE Transactions on Automatic Control*, 48(6), 946–956.
- Lall, S., Krysl, P., & Marsden, J. (2003). Structure-preserving model reduction for mechanical systems. *Physica D*, 184, 304–318.
- Lall, S., Marsden, J. E., & Glavaski, S. (2002). A subspace approach to balanced truncation for model reduction of nonlinear control systems. *International Journal on Robust and Nonlinear Control*, 12, 519–535.
- Le Gall, François (2014). Powers of tensors and fast matrix multiplication. In *International symposium on symbolic and algebraic computation*, Kobe, Japan, July 23–25 (pp. 296–303).
- Lorenz, E. N. (1956). *Empirical orthogonal functions and statistical weather prediction. Scientific report 1, Statistical Forecasting Project*. MIT, Department of Meteorology.
- Majji, M., Juang, J.-N., & Junkins, J. L. (2010). Observer/Kalman-filter time-varying system identification. *Journal of Guidance, Control, and Dynamics*, 33(3), 887–900.
- Mayo, A. J., & Antoulas, A. C. (2007). A framework for the solution of the generalized realization problem. *Linear Algebra and its Applications*, 425(2–3), 634–662.
- Meyer, D. G. (1990). Fractional balanced reduction: model reduction via a fractional representation. *IEEE Transactions on Automatic Control*, 35(12), 1341–1345.
- Moore, B. C. (1981). Principal component analysis in linear systems: controllability, observability, and model reduction. *IEEE Transactions on Automatic Control*, 26(1), 17–32.
- Noack, B. R., Afanasiev, K., Morzynski, M., Tadmor, G., & Thiele, F. (2003). A hierarchy of low-dimensional models for the transient and post-transient cylinder wake. *Journal of Fluid Mechanics*, 497, 335–363.
- Padoan, A., Scarciootti, G., & Astolfi, A. (2016). A geometric characterisation of persistently exciting signals generated by autonomous systems. In *IFAC symposium nonlinear control systems*, Monterey, CA, USA, August 23–25 (pp. 838–843).
- Padoan, A., Scarciootti, G., & Astolfi, A. (2016). A geometric characterisation of the persistence of excitation condition for the solutions of autonomous systems. *IEEE Transactions on Automatic Control*, in press.
- Park, J., & Sandberg, I. W. (1991). Universal approximation using radial-basis-function networks. *Neural Computation*, 3(2), 246–257.
- Rebolho, D. C., Belo, E. M., & Marques, F. D. (2014). Aeroelastic parameter identification in wind tunnel testing via the extended eigensystem realization algorithm. *Journal of Vibration and Control*, 20(11), 1607–1621.
- Rocha, H. (2009). On the selection of the most adequate radial basis function. *Applied Mathematical Modelling*, 33(3), 1573–1583.
- Rodriguez, H., Ortega, R., & Astolfi, A. (2005). Adaptive partial state feedback control of the DC-to-DC Ćuk converter. In *Proceedings of the 2005 American control conference*, 7, June. (pp. 5121–5126).
- Rowley, C. W., Colonius, T., & Murray, R. M. (2004). Model reduction for compressible flows using POD and Galerkin projection. *Physica D: Nonlinear Phenomena*, 189(1–2), 115–129.
- Safonov, M. G., Chiang, R. Y., & Limebeer, D. J. N. (1990). Optimal Hankel model reduction for nonminimal systems. *IEEE Transactions on Automatic Control*, 35(4), 496–502.
- Scarciootti, G. (2015). Model reduction of power systems with preservation of slow and poorly damped modes. In *IEEE power & energy society general meeting*, Denver, Colorado, July 26–30 (pp. 1–5).
- Scarciootti, G. (2017). Low computational complexity model reduction of power systems with preservation of physical characteristics. *IEEE Transactions on Power Systems*, 32(1), 743–752.
- Scarciootti, G., & Astolfi, A. (2015a). Characterization of the moments of a linear system driven by explicit signal generators. In *Proceedings of the 2015 American control conference*, Chicago, IL, July (pp. 589–594).
- Scarciootti, G., & Astolfi, A. (2015b). Model reduction for linear systems and linear time-delay systems from input/output data. In *2015 European control conference*, Linz, July (pp. 334–339).
- Scarciootti, G., & Astolfi, A. (2015c). Model reduction for nonlinear systems and nonlinear time-delay systems from input/output data. In *Proceedings of the 54th IEEE conference on decision and control*, Osaka, Japan, December 15–18 (pp. 7298–7303).
- Scarciootti, G., & Astolfi, A. (2016a). Model reduction by matching the steady-state response of explicit signal generators. *IEEE Transactions on Automatic Control*, 61(7), 1995–2000.
- Scarciootti, G., & Astolfi, A. (2016b). Model reduction of neutral linear and nonlinear time-invariant time-delay systems with discrete and distributed delays. *IEEE Transactions on Automatic Control*, 61(6), 1438–1451.
- Scarciootti, G., Jiang, Z. P., & Astolfi, A. (2016). Constrained optimal reduced-order models from input/output data. In *Proceedings of the 55th IEEE conference on decision and control*, Las Vegas, NV, USA, December 12–14 (pp. 7453–7458).
- Scherpen, J. M. A., & Gray, W. S. (2000). Minimality and local state decompositions of a nonlinear state space realization using energy functions. *IEEE Transactions on Automatic Control*, 45(11), 2079–2086.
- Scherpen, J. M. A., & Gray, W. S. (2002). Nonlinear Hilbert adjoints: Properties and applications to Hankel singular value analysis. *Nonlinear Analysis. Theory, Methods & Applications*, 51(5), 883–901.
- Soberg, J., Fujimoto, K., & Glad, T. (2007). Model reduction of nonlinear differential-algebraic equations. In *IFAC symposium nonlinear control systems*, Pretoria, South Africa, 7 (pp. 712–717).
- Toth, R. (2010). *Lecture notes in control and information sciences, Modeling and identification of linear parameter-varying systems*. Berlin Heidelberg: Springer.
- van Overschee, P., & de Moor, L. R. (1996). *Subspace identification for linear systems: theory, implementation, applications*. Kluwer Academic Publishers.
- Verhaegen, M., & Verdult, V. (2007). *Filtering and system identification: A least squares approach*. Cambridge University Press.
- Verriest, E., & Gray, W. (1998). Dynamics near limit cycles: Model reduction and sensitivity. In *Symposium on mathematical theory of networks and systems*, Padova, Italy.
- Vrabie, D., Pastravanu, O., Abu-Khalaf, M., & Lewis, F. L. (2009). Adaptive optimal control for continuous-time linear systems based on policy iteration. *Automatica*, 45(2), 477–484.
- Wang, Q., & Zhang, L. (2011). Online updating the generalized inverse of centered matrices. In *Proceedings of the 25th AAAI conference on artificial intelligence* (pp. 1826–1827).
- Willcox, K., & Peraire, J. (2002). Balanced model reduction via the proper orthogonal decomposition. *American Institute of Aeronautics and Astronautics*, 40(11), 2323–2330.



Giordano Scarciootti was born in Frascati (Rome), Italy, in 1988. He received his B.S and M.S. degrees in Automation Engineering from the University of Rome "Tor Vergata", Italy, in 2010 and 2012, respectively. In 2012 he joined the Control and Power Group, Imperial College London, London, UK, where he obtained a Ph.D. degree in 2016 with a thesis on approximation, analysis and control of large-scale systems. He is currently a Junior Research Fellow in the Control and Power Group. His research interests are focused on blending control theory and power applications, with particular attention to problems of stability analysis, optimal control and model reduction. He is the recipient of the IET Control & Automation Ph.D. Award (2016).



Alessandro Astolfi was born in Rome, Italy, in 1967. He graduated in electrical engineering from the University of Rome in 1991. In 1992 he joined ETH-Zurich where he obtained a M.Sc. in Information Theory in 1995 and the Ph.D. degree with Medal of Honor in 1995 with a thesis on discontinuous stabilization of nonholonomic systems. In 1996 he was awarded a Ph.D. from the University of Rome "La Sapienza" for his work on nonlinear robust control. Since 1996 he has been with the Electrical and Electronic Engineering Department of Imperial College London, London (UK), where he is currently Professor in Nonlinear Control Theory and Head of the Control and Power Group.

From 1998 to 2003 he was also an Associate Professor at the Dept. of Electronics and Information of the Politecnico di Milano. Since 2005 he has also been a Professor at Dipartimento di Ingegneria Civile e Ingegneria Informatica, University of Rome Tor Vergata. He has been a visiting lecturer in "Nonlinear Control" in several universities, including ETH-Zurich (1995–1996); Terza University of Rome (1996); Rice University, Houston (1999); Kepler University, Linz (2000); SUPELEC, Paris (2001).

His research interests are focused on mathematical control theory and control applications, with special emphasis for the problems of discontinuous stabilization, robust and adaptive control, observer design and model reduction. He is the author of more than 120 journal papers, of 30 book chapters and of over 240 papers in refereed conference proceedings. He is the recipient of the IEEE CSS A. Ruberti Young Researcher Prize (2007) and of the IEEE CSS George S. Axelby Outstanding Paper Award (2012). He is a "Distinguished Member" of the IEEE CSS. He is the author (with D. Karagiannis and R. Ortega) of the monograph "Nonlinear and Adaptive Control with Applications" (Springer-Verlag).

He is Associate Editor of Automatica, the International Journal of Control, the Journal of the Franklin Institute, and the International Journal of Adaptive Control and Signal Processing. He is Senior Editor of the IEEE Trans. on Automatic Control and Editor-in-Chief of the European Journal of Control. He has also served in the IPC of various international conferences. He is currently the Chair of the IEEE CSS Conference Editorial Board.