Moments of random variables: a systems-theoretic interpretation

Alberto Padoan, *Member, IEEE*, and Alessandro Astolfi, *Fellow, IEEE*

***Abstract*— Moments of continuous random variables admitting a probability density function are studied. We show that, under certain assumptions, the moments of a random variable can be characterised in terms of a Sylvester equation and of the steady-state output response of a specific interconnected system. This allows to interpret well-known notions and results of probability theory and statistics in the language of systems theory, including the sum of independent random variables, the notion of mixture distribution and results from renewal theory. The theory developed is based on tools from the center manifold theory, the theory of the steady-state response of nonlinear systems, and the theory of output regulation. Our formalism is illustrated by means of several examples and can be easily adapted to the case of discrete and of multivariate random variables.**

***Index Terms*— Differential equations, interpolation, nonlinear systems, probability, random variables, statistics, transfer functions.**

I. Introduction

One of the most important and challenging problems in probability theory and statistics is that of determining the probability distribution of the random process that is thought to have produced a given set of data. Without making any assumptions on the underlying random process, this problem is extremely difficult in general. In parametric statistics [1 – 4], where the probability distribution which generates the given data is assumed to belong to a family of probability distributions parameterized by a fixed set of parameters, the problem can be dealt with using different approaches. A classical way [1 – 4] to solve this problem is to find mathematical objects that, under specific assumptions, uniquely identify a probability distribution in the family. In probability theory [5 – 8], a similar necessity arises when one is confronted with the problem of specifying uniquely a probability measure through a sequence of numbers. The determination of simple, yet meaningful, objects with these features is therefore of paramount importance. The significance of the moments of a random variable in this context is comparable to that of the derivatives (at a point) for an analytic function. The representation of an analytic function through its Taylor series [9] allows to know the whole function once all the derivatives (at a point) are specified. Similarly, the set of all moments of a random variable uniquely identifies the probability distribution of the random variable, provided that certain conditions are satisfied. This means that the essential features of a random variable can be captured by its moments, which can be then used as an alternative description of the random variable.

Moments have been used in a number of different contexts in probability theory, including the Stieltjes moment problem [10], the Hamburger moment problem [11], the Hausdorff moment problem [12] and the Vorobyev moment problem [13]. Moments have proved instrumental in the first rigorous (yet incomplete) proof of the central limit theorem [14], but also in the proof of the Feynman-Kac formula [15], in the characterisation of the eigenvalues of random matrices [16], and in the study of birth-death processes [17]. The correspondence between moments and probability distributions is also exploited in statistics to fit curves and to design parameter estimation procedures [1 – 4]. For example, the method of moments introduced in [18] takes advantage of this correspondence to build consistent (usually biased) estimators. Approximations of probability density functions may be computed exploiting such correspondence through Pearson's and Johnson's curves [2]. A revisitation of this correspondence has also led to the exploratory projection pursuit [19 – 21], a graphical visualisation method for the interpretation of high-dimensional data. While this list of examples is far from being exhaustive, it illustrates the central role of moments in a number of interesting problems. Further detail on theory and applications of moments can be found, *e.g.*, in [22 – 24].

This work represents a step towards bridging the gap between the notions of moment in probability theory and in systems theory. Connections are first established between moments of random variables and moments of systems [25]. These connections are then used to support the claim that a system can be seen as an *alternative, equivalent* description of the probabilistic structure of a random variable. This, in turn, indicates that systems-theoretic techniques and tools can be used to revisit and shed new light on classical results of probability theory and statistics.

The benefits of a dialogue between the two areas of research has been highlighted, *e.g.*, in [26] and briefly mentioned in [27]. The literature on the analysis of results of probability theory in systems-theoretic terms, however, appears to be limited to specific topics and scattered across different research fields.

This work has been partially supported by the European Union's Horizon 2020 Research and Innovation Programme under grant agreement No 739551 (KIOS CoE).

A. Padoan is with the Department of Engineering, University of Cambridge, Cambridge CB2 1PZ, UK (email: a.padoan@eng.cam.ac.uk).

A. Astolfi is with the Department of Electrical and Electronic Engineering, Imperial College London, London SW7 2AZ, UK and with the Dipartimento di Ingegneria Civile e Ingegneria Informatica, Università di Roma "Tor Vergata", Via del Politecnico 1, 00133, Rome, Italy (email: a.astolfi@imperial.ac.uk).

Linear systems theory has allowed to reinterpret basic results of probability theory for probability density functions with rational Laplace transform, *e.g.*, in [28, 29]. Concepts and tools of systems theory have been used to investigate phase-type distributions [30, 31] and matrix exponential distributions [32]. The study of the properties of these distributions have grown into a well-established area of research which goes under the name of "matrix analytic methods" [33, 34]. Another well-known class of models which has attracted the interest of systems theorists over the past century is that of Markov chains [5, 6, 8, 35 – 37]. One of the reasons for this interest is that these models can be regarded as a special class of positive systems [38], which can be studied exploiting the theory of non-negative matrices [39 – 42]. Queuing systems have been studied borrowing analysis and design tools from systems theory, *e.g.*, in [43, 44], whereas a systems-theoretic approach has been adopted to study risk theory, *e.g.*, in [45, 46]. In circuits theory, moments have been used for the approximation of the propagation delay in a RC tree network by interpreting the impulse response of a circuit as a probability density function [47]. Finally, we emphasise that a number of fundamental tools for modelling, estimation and prediction may not be sharply categorised as belonging to the sphere of probability theory or to that of systems theory, but rather lie at the intersection of these areas of research. The Kalman filter [48, 49] is perhaps the most successful example of exchange of ideas between probability theory and systems theory, but several other examples can be found (see, *e.g.*, [50 – 56] and references therein).

The main objective of this work is to establish a one-to-one correspondence between moments of random variables and moments of systems [25]. This goal is readily achieved by interpreting probability density functions as impulse responses whenever these can be realized by linear time-invariant systems. The situation is more delicate when a probability density function can be only realized, *e.g.*, by means of a linear time-varying system or a system in explicit form [57, 58], for which the interpretation in terms of impulse responses does not possess a direct counterpart. This issue is overcome exploiting recent developments in systems theory [58 – 63], where the moments of a linear system have been characterised as solutions of a Sylvester equation [59, 60] and, under certain hypotheses, as steady-state responses of the output of particular interconnected systems [61 – 63]. The characterisation of the moments of a system in terms of steady-state responses is based on tools arising in the center manifold theory [64], the theory of the steady-state response of a nonlinear system [65], and the output regulation theory for nonlinear systems [66]. Existing results on moments of linear and nonlinear systems [61 – 63] are extended and the notion of moment for systems which only possess a representation in explicit form is revisited [58]. As a direct by-product of our approach, Sylvester equations and Sylvester-like differential equations may be interpreted as powerful computational tools which allow to calculate moments of random variables. Our approach allows to revisit and re-interpret well-known notions and results of probability theory and statistics using the language of systems theory, including the sum of independent random variables [6, 8, 67], the notion of mixture distribution [7, 68], and results from renewal theory [6, 7, 69, 70]. Given the conceptual nature of the paper, we focus on linear systems, linear time-delay systems and systems in explicit form to streamline the exposition, although generalisations to further classes of systems are possible. The formalism is developed for univariate continuous random variables, but can be easily extended to discrete and multivariate random variables.

Preliminary results have been presented in [71], where a first connection between moments of probability density functions and moments of linear time-invariant systems has been established. As a result, the moment generating function of a random variable and the solution of a Sylvester equation have been shown to be closely related. The present manuscript extends our results to probability density functions which may not be described by means of a linear time-invariant system. To this end, the notion of moment of a system is revisited and extended to systems the impulse response of which is defined on the whole real line or on a compact subset of the real line. The theory is supported by several worked-out examples and applications to identifiability, queueing theory and renewal theory.

The rest of the paper is organized as follows. Section II provides basic definitions and the necessary background concerning moments of linear systems, of time-delay systems, of systems in explicit form and of random variables. Section III contains our main results and includes a characterisation of the moments of a random variable admitting a (linear or explicit) realization in terms of the systems-theoretic notion of moment. The moments of probability density functions defined on the real axis as well as probability density functions with compact support are also characterised by means of Sylvester equations and steady-state responses. Section IV provides selected applications of the theoretical results developed, including the sum of independent random variables, the notion of mixture distribution, and results from renewal theory. Finally, conclusions and future research directions are outlined in Section V.

Notation: $\mathbb{Z}_{\geq 0}$ and $\mathbb{Z}_{>0}$ denote the set of non-negative integer numbers and the set of positive integer numbers, respectively. $\mathbb{R}$, $\mathbb{R}^n$ and $\mathbb{R}^{p\times m}$ denote the set of real numbers, of n-dimensional vectors with real entries and of $p \times m$-dimensional matrices with real entries, respectively. $\mathbb{R}_{\geq 0}$ and $\mathbb{R}_{>0}$ denote the set of non-negative real numbers and the set of positive real numbers, respectively. $\mathbb{C}$ denotes the set of complex numbers. $\mathbb{C}_{<0}$ denotes the set of complex numbers with negative real part. i denotes the imaginary unit. I denotes the identity matrix. e_j denotes the vector with the j-th entry equal to one and all other entries equal to zero. $\sigma(A)$ denotes the spectrum of the matrix $A \in \mathbb{R}^{n\times n}$. M' denotes the transpose of the matrix $M \in \mathbb{R}^{p\times m}$. $\|x\|$ denotes the standard Euclidean norm of the vector $x \in \mathbb{R}^n$ while, with some abuse of notation, $\|A\|$ denotes the induced norm of the matrix $A \in \mathbb{R}^{n\times n}$. Δ^k denotes, for all $k \in \mathbb{Z}_{\geq 0}$, the standard k-simplex, *i.e.* $\Delta^k = \left\{ w \in \mathbb{R}^{k+1}_{\geq 0} : w_1 + \ldots + w_{k+1} = 1 \right\}$. $\mathrm{diag}(\lambda)$ denotes the diagonal matrix whose diagonal elements are the entries of the vector $\lambda \in \mathbb{R}^n$. J_0 denotes the matrix

with ones on the superdiagonal and zeros elsewhere. J_λ denotes the Jordan block associated with the eigenvalue $\lambda \in \mathbb{C}$, *i.e.* $J_\lambda = \lambda I + J_0$. $k!$ and $(k)!!$ denote the factorial and the double factorial of $k \in \mathbb{Z}_{\geq 0}$, respectively. $\dot{x}$ denotes the time derivative of the function x, provided it exists. $f_1 * f_2$ denotes the convolution of the functions f_1 and f_2. $\mathcal{L}\{f\}$ denotes the bilateral Laplace transform of the function f. δ_0 and δ_{-1} denote the Dirac δ-function and the right-continuous Heaviside unit step function, respectively. The time reversal of the function $f : \mathbb{R} \to \mathbb{R}$ is defined as $t \mapsto f(-t)$. $\eth_{-1}$ denotes the right-continuous time reversal of δ_{-1}. $\mathbb{1}_{\mathcal{S}}$ denotes the indicator function of the subset $\mathcal{S}$ of a set $\mathcal{X}$, *i.e.* the function $\mathbb{1}_{\mathcal{S}} : \mathcal{X} \to \{0, 1\}$ defined as $\mathbb{1}_{\mathcal{S}}(x) = 1$ if $x \in \mathcal{S}$ and as $\mathbb{1}_{\mathcal{S}}(x) = 0$ if $x \notin \mathcal{S}$. By a convenient abuse of notation, $\mathbb{1}$ also denotes the vector the elements of which are all equal to one. The translation of the function $f : \mathbb{R} \to \mathbb{R}$ by $\tau \in \mathbb{R}$ is defined as $f_\tau : t \mapsto f(t - \tau)$. $\mathcal{C}[-T, 0]$ denotes the set of continuous functions mapping the interval $[-T, 0]$ into $\mathbb{R}^n$ with the topology of uniform convergence. For all $s^\star \in \mathbb{C}$ and $A(s) \in \mathbb{C}^{n\times n}$, $s^\star \in \mathbb{C} \setminus \sigma(A(s))$ denotes $\det(s^\star I - A(s^\star)) \neq 0$, while $\sigma(A(s)) \subset \mathbb{C}_{<0}$ indicates that if $s^\star \in \mathbb{C}$ is such that $\det(s^\star I - A(s^\star)) = 0$, then $s^\star \in \mathbb{C}_{<0}$. $\mathbb{E}[X]$ denotes the expectation of the random variable X.

II. Preliminaries

This section contains a short digression on the notion of moment, which is used with different meanings in the literature and throughout the paper. To give a compact and self-contained exposition, we first provide some material on the notion of moment of a linear system [25, 59–62]. We then illustrate and extend notions and results concerning moments of a system in explicit form [58] and of a time-delay systems [72]. Finally, we recall the notion of moment of a random variable [73].

A. Moments of a linear system

Consider a single-input, single-output, continuous-time, linear, time-invariant system described by the equations

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

in which $x(t) \in \mathbb{R}^n$, $u(t) \in \mathbb{R}$, $y(t) \in \mathbb{R}$ and $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times 1}$ and $C \in \mathbb{R}^{1\times n}$ are constant matrices. Throughout the paper we assume that the system (1) is minimal, *i.e.* reachable and observable, and let $W(s) = C(sI - A)^{-1}B$ be its transfer function.

Definition 1. [61] The moment of order zero of system (1) at $s^\star \in \mathbb{C} \setminus \sigma(A)$ is the complex number $\eta_0(s^\star) = W(s^\star)$. The moment of order $k \in \mathbb{Z}_{>0}$ of system (1) at $s^\star \in \mathbb{C} \setminus \sigma(A)$ is the complex number

$$\eta_k(s^\star) = \frac{(-1)^k}{k!} \left[\frac{d^k}{ds^k} W(s) \right]_{s=s^\star}. \tag{2}$$

The moments of system (1) can be characterised in terms of the solution of a Sylvester equation [59–62]. The following statements provide a version of the results presented in [61, 62].

Lemma 1. Consider system (1). Assume $k \in \mathbb{Z}_{\geq 0}$ and $s^\star \in \mathbb{C} \setminus \sigma(A)$. Then $\begin{bmatrix} \eta_k(s^\star) & \cdots & \eta_1(s^\star) & \eta_0(s^\star) \end{bmatrix}' = \Psi_k \Upsilon_k B$, where $\Psi_k \in \mathbb{R}^{(k+1)\times(k+1)}$ is a signature matrix[1] and $\Upsilon_k \in \mathbb{R}^{(k+1)\times n}$ is the unique solution of the Sylvester equation

$$J_{s^\star} \Upsilon_k + e_{k+1} C = \Upsilon_k A. \tag{3}$$

Lemma 2. Consider system (1). Assume $k \in \mathbb{Z}_{\geq 0}$ and $s^\star \in \mathbb{C} \setminus \sigma(A)$. Then there exists a one-to-one correspondence between the moments $\eta_0(s^\star), \eta_1(s^\star), \ldots, \eta_k(s^\star)$ and the matrix $\Upsilon_k B$, where $\Upsilon_k \in \mathbb{R}^{(k+1)\times n}$ is the unique solution of the Sylvester equation

$$S\Upsilon_k + MC = \Upsilon_k A, \tag{4}$$

in which $S \in \mathbb{R}^{(k+1)\times(k+1)}$ is a non-derogatory[2] matrix with characteristic polynomial

$$\chi(s) = (s - s^\star)^{k+1} \tag{5}$$

and $M \in \mathbb{R}^{k+1}$ is such that the pair (S, M) is reachable.

The moments of system (1) can be also described under special circumstances by means of the (well-defined) steady-state output response[3] of the interconnection of system (1) with a system described by equations

$$\dot{\omega} = S\omega + Mv, \quad d = \omega, \tag{6}$$

with $\omega(t) \in \mathbb{R}^{k+1}$, $v(t) \in \mathbb{R}$, $d(t) \in \mathbb{R}^{k+1}$ and $v = y$, *i.e.* of the interconnected system

$$\dot{x} = Ax + Bu, \quad \dot{\omega} = S\omega + MCx, \quad d = \omega, \tag{7}$$

in which $S \in \mathbb{R}^{(k+1)\times(k+1)}$ is a non-derogatory matrix with characteristic polynomial (5), $M \in \mathbb{R}^{k+1}$ is such that the pair (S, M) is reachable, as detailed by the following statement, which is a variation of [62, Theorem 1]. In what follows, we tacitly assume that the pair (S, M) always satisfies these properties.

Theorem 1. Consider system (1), system (6), and the interconnected system (7). Assume $\sigma(A) \subset \mathbb{C}_{<0}$, $s^\star \in \mathbb{C} \setminus \sigma(A)$, $x(0) = 0$, $\omega(0) = 0$ and $u = \delta_0$. Then there exists a one-to-one correspondence[4] between the moments $\eta_0(s^\star), \eta_1(s^\star), \ldots, \eta_k(s^\star)$ and the (well-defined) steady-state response of the output d of system (7).

B. Moments of a time-delay system

Consider a single-input, single-output, continuous-time, linear, time-delay system with discrete constant delays described by equations

$$\dot{x}(t) = \sum_{j=0}^{\varsigma} A_j x_{\tau_j} + \sum_{j=\varsigma+1}^{\varrho} B_j u_{\tau_j}, \; y(t) = \sum_{j=0}^{\varsigma} C_j x_{\tau_j}, \tag{8}$$

with $x(\theta) = \varphi(\theta)$ for $-T \leq \theta \leq 0$, in which $x(t) \in \mathbb{R}^n$, $u(t) \in \mathbb{R}$, $y(t) \in \mathbb{R}$, $\tau_0 = 0$, $\tau_j \in \mathbb{R}_{>0}$, for $1 \leq j \leq \varrho$,

[1] A signature matrix is a diagonal matrix with ± 1 on the diagonal [74].

[2] A matrix is non-derogatory if its characteristic and minimal polynomials coincide [41].

[3] The notion of steady-state response is taken from [75] (see also [65, 66, 76]). Note that if system (1) is stable and the signal u is bounded backward and forward in time, system (1) has a unique, well-defined steady-state response.

[4] The terminology is borrowed from [61, 62]. By one-to-one correspondence we mean that the steady-state response of the output d is uniquely determined by the moments and *vice versa*.

$T = \max_{0\le j\le \varrho}\tau_j$, $\varphi \in \mathcal{C}[-T,0]$, and $A_j \in \mathbb{R}^{n\times n}$, $B_j \in \mathbb{R}^{n\times 1}$ and $C_j \in \mathbb{R}^{1\times n}$ are constant matrices for $0 \le j \le \varsigma$ and $\varsigma+1 \le j \le \varrho$, respectively. Throughout the paper the system (8) is assumed to be minimal, *i.e.* reachable and observable, and let $W(s) = C(s)(sI - A(s))^{-1}B(s)$ be its transfer function, with $A(s) = \sum_{j=0}^{\varsigma} A_j e^{-\tau_j s}$, $B(s) = \sum_{j=\varsigma+1}^{\varrho} B_j e^{-\tau_j s}$ and $C(s) = \sum_{j=0}^{\varsigma} C_j e^{-\tau_j s}$.

Definition 2. [72] The moment of order zero of system (8) at $s^\star \in \mathbb{C}\setminus\sigma(A(s))$ is the complex number $\eta_0(s^\star) = W(s^\star)$. The moment of order $k \in \mathbb{Z}_{>0}$ of system (8) at $s^\star \in \mathbb{C}\setminus\sigma(A(s))$ is the complex number

$$\eta_k(s^\star) = \frac{(-1)^k}{k!}\left[\frac{d^k}{ds^k}W(s)\right]_{s=s^\star}. \tag{9}$$

A theory of moments can be also developed for time-delay systems exploiting a particular Sylvester-like equation [72]. For completeness, we briefly formulate the results of [72] in a form that is more convenient for our purposes.

Lemma 3. Consider system (8). Assume $k \in \mathbb{Z}_{\ge 0}$ and $s^\star \in \mathbb{C}\setminus\sigma(A(s))$. Then $[\eta_k(s^\star)\ \cdots\ \eta_1(s^\star)\ \eta_0(s^\star)]' = \sum_{j=\varsigma+1}^{\varrho}\Psi_k e^{-J_{s^\star}\tau_j}\Upsilon_k B_j$, where $\Psi_k \in \mathbb{R}^{(k+1)\times(k+1)}$ is a signature matrix and $\Upsilon_k \in \mathbb{R}^{(k+1)\times n}$ is the unique solution of the equation

$$J_{s^\star}\Upsilon_k + \sum_{j=0}^{\varsigma} e_{k+1}C_j e^{-J_{s^\star}\tau_j} = \sum_{j=0}^{\varsigma} A_j\Upsilon_k e^{-J_{s^\star}\tau_j}. \tag{10}$$

Lemma 4. Consider system (8). Assume $k \in \mathbb{Z}_{\ge 0}$ and $s^\star \in \mathbb{C}\setminus\sigma(A(s))$. Then there exists a one-to-one correspondence between the moments $\eta_0(s^\star), \eta_1(s^\star), \dots, \eta_k(s^\star)$ and the matrix $\sum_{j=\varsigma+1}^{\varrho} e^{-S\tau_j}\Upsilon_k B_j$, where $\Upsilon_k \in \mathbb{R}^{(k+1)\times n}$ is the unique solution of the equation

$$S\Upsilon_k + \sum_{j=0}^{\varsigma} MC_j e^{-S\tau_j} = \sum_{j=0}^{\varsigma} A_j\Upsilon_k e^{-S\tau_j}, \tag{11}$$

in which $S \in \mathbb{R}^{(k+1)\times(k+1)}$ is a non-derogatory matrix with characteristic polynomial (5) and $M \in \mathbb{R}^{k+1}$ is such that the pair (S, M) is reachable.

C. Moments of a system in explicit form

Consider a single-output, continuous-time system in explicit form[5] described by equations

$$x(t) = \Lambda(t,t_0)x_0, \quad y(t) = Cx(t), \tag{12}$$

in which $x(t) \in \mathbb{R}^n$, $y(t) \in \mathbb{R}$, $t_0 \in \mathbb{R}_{\ge 0}$, $x(t_0) = x_0 \in \mathbb{R}^n$, $C \in \mathbb{R}^{1\times n}$ and $\Lambda : \mathbb{R}\times\mathbb{R} \to \mathbb{R}^{n\times n}$ is piecewise continuously differentiable in the first argument and such that $\Lambda(t_0,t_0) = I$.

In analogy with the theory developed for linear systems, a characterisation of the steady-state output response of the interconnection of system (6) and of system (12), with $v = y$, *i.e.* of the system

$$x(t) = \Lambda(t,t_0)x_0, \quad \dot\omega = S\omega + MCx, \quad d = \omega, \tag{13}$$

can be given in terms of the solution of a matrix differential equation which plays the role of the Sylvester equation (4). This, in turn, allows to define a notion of moment for systems in explicit form [77]. To this end, the following technical assumptions are needed.

Assumption 1. The matrix $\Lambda(t,t_0)$ is non-singular for every $t \ge t_0$.

Assumption 2. The function $t \mapsto \dot\Lambda(t,t_0)\Lambda(t,t_0)^{-1}$ is piecewise continuous.

Assumption 3. There exist $K \in \mathbb{R}_{\ge 0}$ and $\alpha \in \mathbb{R}_{>0}$ such that $\|\Lambda(t,t_0)\| \le Ke^{-\alpha(t-t_0)}$ for every $t \ge t_0$.

Assumption 4. The point $s^\star \in \mathbb{C}$ is such that $\mathrm{Re}(s^\star) < -\alpha$.

Assumption 5. The entries of Λ have a strictly proper Laplace transform.

Assumption 6. The pair (x_0, C) is such that the poles of the Laurent series associated with $\mathcal{L}\{C\Lambda x_0\}$ and $\mathcal{L}\{\Lambda\}$ coincide.

Remark 1. The manifold

$$\mathcal{M} = \left\{(x,\omega,t) \in \mathbb{R}^{n+k+2} : \omega(t) = \Upsilon(t)x(t)\right\}, \tag{14}$$

is an invariant integral manifold[6] of system (13) whenever the function $\Upsilon : \mathbb{R} \to \mathbb{R}^{(k+1)\times n}$ satisfies (except at points of discontinuity) the ordinary differential equation

$$\dot\Upsilon(t) = S\Upsilon(t) - \Upsilon(t)\dot\Lambda(t,t_0)\Lambda(t,t_0)^{-1} + MC, \tag{15}$$

the solution of which is

$$\Upsilon(t) = e^{S(t-t_0)}\Upsilon(t_0)\Lambda(t_0,t) + \int_{t_0}^{t} e^{S(t-\zeta)}MC\Lambda(\zeta,t)d\zeta, \tag{16}$$

with initial condition $\Upsilon(t_0) \in \mathbb{R}^{(k+1)\times n}$, for all $t \ge t_0$. △

The following result is instrumental to define the notion of moment for system (12) and is adapted to the purposes of this paper from a statement originally presented in [77].

Theorem 2. Consider system (6), system (12) and the interconnected system (13). Suppose Assumptions 1, 2, 3 and 4 hold. Then there exists a unique matrix $\Upsilon_\infty(t_0) \in \mathbb{R}^{(k+1)\times n}$ such that $\lim_{t\to\infty}\|\Upsilon(t) - \Upsilon_\infty(t)\| = 0$ for any $\Upsilon(t_0) \in \mathbb{R}^{(k+1)\times n}$, where Υ and Υ_∞ are the solutions of (15) with initial conditions $\Upsilon(t_0)$ and $\Upsilon_\infty(t_0)$, respectively. Moreover, the manifold (14) is an attractive invariant integral manifold of system (13).

Proof. To begin with note that by Assumption 1 the right-hand side of (15) is well-defined. Let Υ_1 and Υ_2 be the solutions of (15) (except at point of discontinuity) corresponding to the initial conditions $\Upsilon_1(t_0) \in \mathbb{R}^{(k+1)\times n}$ and $\Upsilon_2(t_0) \in \mathbb{R}^{(k+1)\times n}$. Defining the function E as the difference of Υ_1 and Υ_2 and differentiating with respect to the time argument yields the ordinary differential equation $\dot E(t) = SE(t) - E(t)\dot\Lambda(t,t_0)\Lambda(t,t_0)^{-1}$ the solution of which, in view of (16), is $E(t) = e^{S(t-t_0)}E(t_0)\Lambda(t_0,t)$. By Assumptions 3 and 4, this implies $\lim_{t\to\infty}\|E(t)\| = 0$. As a result, there exists a solution Υ_∞ to which every solution of (15) converges asymptotically, *i.e.* there exists $\Upsilon_\infty(t_0) \in \mathbb{R}^{(k+1)\times n}$ such that $\lim_{t\to\infty}\|\Upsilon(t) - \Upsilon_\infty(t)\| = 0$ for any $\Upsilon(t_0) \in \mathbb{R}^{(k+1)\times n}$, where Υ and Υ_∞ are the solutions of (15) with initial conditions $\Upsilon(t_0)$ and $\Upsilon_\infty(t_0)$, respectively.

[5] The terminology is borrowed from [57, 58].

[6] See [78].

Moreover, by Assumption 2, Υ_∞ is unique. To complete the proof, note that by Remark 1 the set (14) is an invariant integral manifold of system (13) and that the set (14) is attractive since every solution of (15) converges asymptotically to Υ_∞. □

Theorem 2 allows to introduce the following definition, which is reminiscent of the notion of moment given in [58].

Definition 3. Consider system (6), system (12) and the interconnected system (13). Suppose Assumptions 1, 2, 3 and 4 hold. The moment of system (12) at $s^\star$ is the function $\Upsilon_\infty x$, where Υ_∞ is the unique solution of (15) with $\Upsilon(t_0) = 0$.

Theorem 3. Consider system (6), system (12) and the interconnected system (13). Suppose Assumptions 1, 2, 3, 4, 5 and 6 hold. Then there exists a one-to-one correspondence between the moment of system (12) at $s^\star$ and the (well-defined) steady-state response of the output d of system (13).

Proof. By Assumptions 1, 2, 3 and 4, in view of Theorem 2, the steady-state response of the output d is well-defined. By Assumption 5 and 6, the steady-state response of the output d coincides with $\Upsilon_\infty x$, which by definition is the moment of system (12) at $s^\star$. □

Remark 2. Equations (12) describe a considerably general class of continuous-time signals. In particular, under the stated assumptions, this class contains all (exponentially bounded) piecewise continuously differentiable functions that can be generated as the solution of a single-input, single-output, continuous-time, linear, time-varying system of the form

$$\dot{x} = A(t)x + B(t)u, \quad y = Cx, \tag{17}$$

with $x(t) \in \mathbb{R}^n$, $u(t) \in \mathbb{R}$, $y(t) \in \mathbb{R}$, $x(0) = 0$, $u = \delta_0$, $A : \mathbb{R} \to \mathbb{R}^{n\times n}$ defined as $A(t) = \dot{\Lambda}(t, t_0)\Lambda(t, t_0)^{-1}$ for all $t \geq t_0$ such that Λ is differentiable in the first argument and $B : \mathbb{R} \to \mathbb{R}^n$ such that the pair $(A(t), B(t))$ is reachable for all $t \in \mathbb{R}_{\geq 0}$. △

D. Moments of a random variable

For completeness, we now recall the notion of moment of a random variable.

Definition 4. [73] The moment of order $k \in \mathbb{Z}_{\geq 0}$ of the random variable X is defined as $\mu_k = \mathbb{E}[X^k]$ whenever the expectation exists.

To simplify the exposition, in the sequel we consider exclusively continuous random variables admitting a probability density function with finite moments of all orders. We also ignore all measure-theoretic considerations as they are not essential for any of the arguments. A discussion on the extension of our results to more general situations is deferred to Section V.

To illustrate our results and to demonstrate our approach we use several worked-out examples throughout this work.

Example 1 (The exponential distribution). The probability density function of a random variable X having exponential distribution with parameter $\lambda \in \mathbb{R}_{>0}$ is defined as

$$f_X : \mathbb{R} \to \mathbb{R},\ t \mapsto \lambda e^{-\lambda t}\delta_{-1}(t). \tag{18}$$

A direct computation shows that the moment of order $k \in \mathbb{Z}_{>0}$ of the random variable X satisfies the relation $\mu_k = \frac{k}{\lambda}\mu_{k-1}$, with $\mu_0 = 1$, and hence $\mu_k = \frac{k!}{\lambda^k}$. ▲

Example 2 (The hyper-exponential distribution). The probability density function of a random variable X having hyper-exponential distribution is defined as

$$f_X : \mathbb{R} \to \mathbb{R},\ t \mapsto \sum_{j=1}^{n} p_j \lambda_j e^{-\lambda_j t}\delta_{-1}(t), \tag{19}$$

where $n \in \mathbb{Z}_{>0}$ is referred to as the number of phases of X and $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_n) \in \mathbb{R}^n_{>0}$ and $p = (p_1, p_2, \ldots, p_n) \in \Delta^{n-1}$ are given parameters. Observe that f_X can be written as $f_X = \sum_{j=1}^n p_j f_{X_j}$, in which $f_{X_j} : \mathbb{R} \to \mathbb{R},\ t \mapsto \lambda_j e^{-\lambda_j t}\delta_{-1}(t)$ is the probability density function of a random variable X_j having exponential distribution with parameter λ_j. Thus, by linearity of the expectation operator, the moment of order $k \in \mathbb{Z}_{\geq 0}$ of the random variable X is $\mu_k = \sum_{j=1}^n p_j \frac{k!}{\lambda_j^k}$. ▲

III. Main results

This section contains the main results of the paper. The moments of a random variable are shown to be in one-to-one correspondence with the moments of a system at zero. This is established first for the special situation in which a given probability density functions can be identified with the impulse response of a linear system. The theory developed is then extended to the broader class of probability density functions which can be represented by a system in explicit form using the theory of moments presented in the previous section. Finally, the case of probability density functions defined on the whole real axis and on compact sets is considered.

A. Probability density functions realized by linear systems

Definition 5. Consider system (1) and a random variable X with probability density function f_X. The probability density function f_X is realized by system (1) if $f_X(t) = Ce^{At}B\delta_{-1}(t)$ for all $t \in \mathbb{R}$, in which case system (1) is referred to as a (linear) realization of f_X.

Necessary and sufficient conditions for a probability density function to be realized by a linear system can be established using well-known results of linear realization theory [79 – 81]. Note that every linear realization of a probability function must be stable[7], as detailed by the following statement.

Lemma 5. Consider system (1) and a random variable X with probability density function f_X. If system (1) is a realization of f_X, then $\sigma(A) \subset \mathbb{C}_{<0}$.

As a direct application of Definitions 1, 4, 5 and of Lemma 5 the moments of a random variable can be characterised by means of moments of systems.

Theorem 4. Consider system (1) and a random variable X with probability density function f_X. Assume system (1) is a

[7] See [25] for the definition.

realization of f_X. Then the moments of the random variable X and the moments of system (1) at zero are such that

$$\frac{\mu_k}{k!} = \eta_k(0), \tag{20}$$

for all $k \in \mathbb{Z}_{\geq 0}$.

Corollary 1. Consider system (1) and a random variable X with probability density function f_X. Assume system (1) is a realization of f_X. Then the moments up to the order $k \in \mathbb{Z}_{\geq 0}$ of the random variable X are given by the entries of $\Psi_k \Upsilon_k B$, where Υ_k is the unique solution of the Sylvester equation (3), with $\Psi_k = \operatorname{diag}((-1)^k k!, \dots, -1, 1)$ and $s^\star = 0$.

Remark 3. The moments of a random variable can be determined by direct application of Definition 4 or by "pattern matching" using existing tables of moments. The one-to-one correspondence established in Corollary 1, on the other hand, indicates that a closed-form expression for the moments of a random variable can be computed from the solution of a Sylvester equation, which can be solved with numerically reliable techniques [25]. The computation of moments of random variables through Sylvester-like equations is one of the leitmotifs underlying our approach. △

Corollary 2. Consider system (1) and a random variable X with probability density function f_X. Assume system (1) is a realization of f_X. Then the moments up to the order $k \in \mathbb{Z}_{\geq 0}$ of the random variable X are in one-to-one correspondence with the matrix $\Upsilon_k B$, where Υ_k is the unique solution of the Sylvester equation (4), in which S is a non-derogatory matrix with characteristic polynomial (5) and M is such that the pair (S, M) is reachable.

Corollary 3. Consider system (1), system (6), and the interconnected system (7). Let X be a random variable with probability density function f_X. Assume system (1) is a realization of f_X, $s^\star = 0$, $x(0) = 0$, $\omega(0) = 0$ and $u = \delta_0$. Then there exists a one-to-one correspondence between the moments up to the order $k \in \mathbb{Z}_{\geq 0}$ of the random variable X and the (well-defined) steady-state response of the output d of system (7).

Example 3 (The exponential distribution, continued). Consider a random variable X having exponential distribution with probability density function f_X and parameter $\lambda \in \mathbb{R}_{>0}$. A direct inspection shows that the probability density function f_X is realized by the linear, time-invariant system

$$\dot{x} = -\lambda x + \lambda u, \quad y = x, \tag{21}$$

i.e. by system (1) with $A = -\lambda$, $B = \lambda$ and $C = 1$. Note that the only eigenvalue of system (21) has negative real part, which is consistent with Lemma 5. A direct application of Definition 1 yields $\eta_k(0) = \frac{1}{\lambda^k}$, which, in view of Example 1 and in agreement with Theorem 4, shows that the moments of the random variable X are in one-to-one correspondence with the moments of system (21) at zero. In accordance with Corollary 1, setting $\Psi_k = \operatorname{diag}((-1)^{k+1} k!, \dots, -1, 1)$ and $\Sigma_k = J_0$ yields $\Psi_k \Upsilon_k B = \begin{bmatrix} \frac{k!}{\lambda^k} & \cdots & \frac{1}{\lambda} & 1 \end{bmatrix}'$, in which Υ_k is the unique solution of the Sylvester equation (3). By Corollary 2, a one-to-one correspondence can be also inferred between the moments of the random variable X and the Sylvester equation (4). Finally, note that the components of the (well-defined) steady-state response of the output d of system (7) can be written as $d_l(t) = \sum_{j=0}^{l-1} (-1)^j \frac{1}{\lambda^j} t^j = \sum_{j=0}^{l-1} \mu_j \frac{(-t)^j}{j!}$, for all $l \in \{1, 2, \dots, k+1\}$, and hence there is a one-to-one correspondence between the moments up to the order k of the random variable X and the steady-state response of the output d of system (7). ▲

B. Probability density functions realized by systems in explicit form

We have seen that a systems-theoretic interpretation can be given to probability density function which can be realized by a linear system. However, the vast majority of probability density functions cannot be described by a linear time-invariant differential equation. To provide a generalisation of the results established which accounts for more general probability density functions, we develop a parallel of the formulation presented in the previous section using the theory of moments for systems in explicit form [58]. To begin with, we introduce the following definition.

Definition 6. Consider system (12) and a random variable X with probability density function f_X. The probability density function f_X is realized[8] by system (12) if $f_X(t) = C\Lambda(t, t_0) x_0 \delta_{-1}(t)$ for all $t \geq t_0$, in which case system (12) is referred to as a (explicit) realization of f_X.

Theorem 5. Consider system (6), system (12) and the interconnected system (13). Suppose Assumptions 1, 2, 3 and 4 hold. Let X be a random variable with probability density function f_X and assume system (12) is a realization of f_X. Then the moments of the random variable X up to the order $k \in \mathbb{Z}_{\geq 0}$ are in one-to-one correspondence with the moment of system (12) at zero.

Proof. To begin with, note that by Assumptions 1, 2, 3 and 4 the moment of system (12) at zero is well-defined. By Definition 3 and by (16), the moment of system (12) at zero reads as

$$\begin{aligned} \Upsilon_\infty(t) x(t) &= \left(\int_{t_0}^{t} e^{S(t-\zeta)} M C \Lambda(\zeta, t) d\zeta \right) \Lambda(t, t_0) x_0 \\ &= e^{St} \int_{t_0}^{t} e^{-S\zeta} M C \Lambda(\zeta, t_0) x_0 d\zeta \\ &= e^{St} \int_{t_0}^{t} e^{-S\zeta} M f_X(\zeta) d\zeta, \end{aligned}$$

where the last identity holds since system (12) is a realization of f_X. Define $H_+(t) = \int_{t_0}^{t} e^{-S\zeta} M f_X(\zeta) d\zeta$. Since S is a non-derogatory matrix with characteristic polynomial (5), with $s^\star = 0$, and since the pair (S, M) is reachable, there exists a non-singular matrix $T \in \mathbb{R}^{(k+1)\times(k+1)}$ such that $T^{-1} M = e_{k+1}$ and $T^{-1} S T = J_0$. This implies

$$\lim_{t \to \infty} H_+(t) = \begin{bmatrix} \frac{(-1)^k}{k!} \mu_k & \cdots & -\mu_1 & \mu_0 \end{bmatrix}' \tag{22}$$

[8]The impulse response may depend on the time of the impulse t_0 for time varying systems. Note, however, that our purpose is to *model* the probabilistic structure of a random variable representing its probability density function by means of a system and its impulse response. This means that t_0 can be considered as a parameter that can be assigned.

and hence that the components of the moment of system (12) at zero grow polynomially as $t \to \infty$, with coefficients uniquely determined by the moments $\mu_0, \mu_1, \dots, \mu_k$, which proves the claim. □

Remark 4. Assumption 1 is violated by any explicit realization of the form (12) associated with a probability density function with compact support, *i.e.* zero everywhere except on a compact subset of the real line. A discussion on the extension of Theorem 5 to such class of probability density functions is deferred to Section III-D. △

Remark 5. While every linear realization of a probability density function must be internally stable (Lemma 5), it is not possible to prove that every explicit realization of a probability density function must satisfy Assumption 3. The reason is that there exist probability density functions with a "tail" that is "heavier" than the one of the exponential [69, 82], including those of Pareto, Weibull, and Cauchy random variables. Assumption 3 is therefore a strong assumption which rules out important probability density functions. Note, however, that a generalisation of our results to probability density functions with a heavy tail can be established with more advanced measure-theoretic tools. △

Example 4 (The half-normal distribution). The probability density function of a random variable X having half-normal distribution with parameter $\sigma \in \mathbb{R}_{>0}$ is defined as

$$f_X : \mathbb{R} \to \mathbb{R},\ t \mapsto \sqrt{\frac{2}{\pi\sigma^2}} e^{-\frac{t^2}{2\sigma^2}} \delta_{-1}(t). \tag{23}$$

A direct inspection shows that the probability density function f_X is realized by the linear, time-varying system

$$\dot{x} = -\frac{t}{\sigma^2} x + \sqrt{\frac{2}{\pi\sigma^2}} u, \quad y = x, \tag{24}$$

in which $x(t) \in \mathbb{R}$, $u(t) \in \mathbb{R}$, $y(t) \in \mathbb{R}$. Consider now the interconnection of system (6) and of system (24), with $v = y$, set $s^\star = 0$ and note that Assumptions 1, 2, 3 and 4 hold. Equation (15) boils down to

$$\dot{\Upsilon}(t) = \left(S + \frac{t}{\sigma^2} I \right) \Upsilon(t) + M \tag{25}$$

and can be solved by direct application of formula (16), with $\Upsilon(t_0) = 0$, yielding $\Upsilon_\infty(t) = e^{St} H_+(t)$, with

$$H_+(t) = \begin{bmatrix} \frac{(-1)^k}{k!} \int_0^t \zeta^k e^{-\frac{(\zeta^2 - t^2)}{2\sigma^2}} d\zeta \\ \vdots \\ -\int_0^t \zeta e^{-\frac{(\zeta^2 - t^2)}{2\sigma^2}} d\zeta \\ \int_0^t e^{-\frac{(\zeta^2 - t^2)}{2\sigma^2}} d\zeta \end{bmatrix}. \tag{26}$$

Since $x(t) = \sqrt{\frac{2}{\pi\sigma^2}} e^{-\frac{t^2}{2\sigma^2}} \delta_{-1}(t)$, by Definitions 3 and 4, this implies (22) holds, with $\mu_0 = 1$ and

$$\mu_k = \begin{cases} \sigma^k (k-1)!!, & \text{if } k \text{ is even}, \\ \sqrt{\frac{2}{\pi}} \sigma^k (k-1)!!, & \text{if } k \text{ is odd}. \end{cases}$$

In accordance with Theorem 4, this shows that the moments of the random variable X up to the order $k \in \mathbb{Z}_{\geq 0}$ uniquely specify the moment of system (24) at zero as $t \to \infty$. ▲

Corollary 4. Consider system (6), system (12) and the interconnected system (13). Suppose system (12) is a realization of f_X and Assumptions 1, 2, 3, 4, 5 and 6 hold. Then there exists a one-to-one correspondence between the moments up to the order $k \in \mathbb{Z}_{\geq 0}$ of the random variable X and the (well-defined) steady-state response of the output d of system (13).

C. Probability density functions on the whole real axis

The results established so far characterise probability density functions with support on the non-negative real axis. These results are not satisfactory because most probability density functions are defined over the whole real line. This issue, however, can be easily resolved using the following approach.

Every probability density function f_X can be decomposed as the sum of a function f_c which vanishes on the non-positive real axis and of a function f_{ac} which vanishes on the non-negative real axis, *i.e.* $f_X(t) = f_c(t)\delta_{-1}(t) + f_{ac}(t)\delta_{-1}(t)$ for all $t \in \mathbb{R}$. We call f_c the *causal part* of f_X and f_{ac} the *anticausal part* of f_X. Note that the function need not to be continuous, but only integrable.

With these premises, the following result holds.

Theorem 6. Consider a random variable X with probability density function f_X. Let f_c and f_{ac} be the causal part and the anti-causal part of f_X, respectively. Assume f_c is realized by the minimal system

$$\dot{x}_c = A_c x_c + B_c u_c, \quad y_c = C_c x_c, \tag{27}$$

and the time reversal of f_{ac} is realized by the minimal system

$$\dot{x}_{ac} = A_{ac} x_{ac} + B_{ac} u_{ac}, \quad y_{ac} = C_{ac} x_{ac}, \tag{28}$$

with $x_j(t) \in \mathbb{R}^{n_j}$, $u_j(t) \in \mathbb{R}$, $y_j(t) \in \mathbb{R}$, and $A_j \in \mathbb{R}^{n_j \times n_j}$, $B_j \in \mathbb{R}^{n_j \times 1}$ and $C_j \in \mathbb{R}^{1 \times n_j}$ constant matrices for $j \in \{c, ac\}$. Then the moments of the random variable X and the moments of systems (27) and (28) at zero satisfy the identity

$$\frac{\mu_k}{k!} = (-1)^k \eta_k^{ac}(0) + \eta_k^c(0). \tag{29}$$

for every $k \in \mathbb{Z}_{\geq 0}$.

Proof. Let $k \in \mathbb{Z}_{\geq 0}$ and note that

$$\begin{aligned} \mu_k &= \int_{-\infty}^{\infty} t^k f_X(t) dt \\ &= \int_{-\infty}^{0} t^k f_{ac}(t) dt + \int_0^{\infty} t^k f_c(t) dt \\ &= (-1)^k \int_0^{\infty} t^k f_{ac}(-t) dt + \int_0^{\infty} t^k f_c(t) dt \\ &= (-1)^k k! \eta_k^{ac}(0) + k! \eta_k^c(0), \end{aligned}$$

which proves the claim. □

Corollary 5. Consider system (1) and a random variable X with probability density function f_X. Assume f_X is even and the causal part of f_X is realized by system (1). Then the moments of the random variable X and the moments of system (1) at zero satisfy the identity

$$\frac{\mu_k}{k!} = \frac{(-1)^k + 1}{2} \eta_k(0). \tag{30}$$

for all $k \in \mathbb{Z}_{\geq 0}$.

Proof. Let f_c and f_{ac} be the causal part and the anti-causal part of f_X, respectively. By hypothesis the identity $f_X(t) = f_c(t)\delta_{-1}(t) + f_{ac}(t)\mathfrak{d}_{-1}(t) = f_c(t)\delta_{-1}(t) + f_c(-t)\mathfrak{d}_{-1}(t)$ holds for all $t \in \mathbb{R}$. By the time-reversal property of the Laplace transform [83, p. 687], this implies $\mathcal{L}\{f_X\}(s) = \mathcal{L}\{f_c\}(s) + \mathcal{L}\{f_c\}(-s)$, from which the claim follows. □

Example 5 (The Laplace distribution). The probability density function of a random variable X having a Laplace distribution with parameter $\lambda \in \mathbb{R}_{>0}$ is defined as

$$f_X : \mathbb{R} \to \mathbb{R},\ t \mapsto \frac{\lambda}{2} e^{-\lambda|t|}. \tag{31}$$

The causal part of f_X is $f_c : \mathbb{R} \to \mathbb{R}$, $t \mapsto \frac{\lambda}{2} e^{-\lambda t}\delta_{-1}(t)$, while the anticausal part of f_X is $f_{ac} : \mathbb{R} \to \mathbb{R}$, $t \mapsto \frac{\lambda}{2} e^{\lambda t}\mathfrak{d}_{-1}(t)$. The causal part and the time reversal of the anti-causal part of f_X are both realized by the minimal system

$$\dot{x} = -\lambda x + \frac{\lambda}{2} u, \quad y = x, \tag{32}$$

in which $x(t) \in \mathbb{R}$, $u(t) \in \mathbb{R}$, $y(t) \in \mathbb{R}$. Thus, by Theorem 6, the moment of order $k \in \mathbb{Z}_{\geq 0}$ of the random variable X is

$$\mu_k = (-1)^k \frac{k!}{2\lambda^k} + \frac{k!}{2\lambda^k}. \tag{33}$$

In agreement with Corollary 5, since f_X is even and the causal part of f_X is realized by system (32), the moment of order $k \in \mathbb{Z}_{\geq 0}$ of the random variable X can be written as $\mu_k = \frac{(-1)^k + 1}{2} \frac{k!}{\lambda^k}$, which is consistent with formula (33). Finally, we emphasise that a simple exercise in integration shows that the moments of the random variable X are indeed given by (33). ▲

Remark 6. Theorem 6 and Corollary 5 allow to establish Corollaries 1, 2 and 3 for a random variable X with probability density function defined on the whole real axis, provided that its causal part and the time reversal of its anti-causal part are realized by systems of the form (27) and (28), respectively. This can be achieved noting that the moments up to the order $k \in \mathbb{Z}_{\geq 0}$ of the random variable X are given by the entries of $\Psi_T \Upsilon_T B_T$, where $\Psi_T = \begin{bmatrix} D_k & \Psi_k D_k \end{bmatrix} \in \mathbb{R}^{(k+1)\times(2k+2)}$, with $\Psi_k \in \mathbb{R}^{(k+1)\times(k+1)}$ a signature matrix and $D_k = \operatorname{diag}(k!, \ldots, 1!, 1)$, and $\Upsilon_T \in \mathbb{R}^{(2k+2)\times(n_a+n_{ac})}$ is the unique solution of the Sylvester equation

$$J_0 \Upsilon_T + e_{k+1} C_T = \Upsilon_T A_T, \tag{34}$$

with

$$A_T = \begin{bmatrix} A_{ac} & 0 \\ 0 & A_c \end{bmatrix},\ B_T = \begin{bmatrix} B_{ac} \\ B_c \end{bmatrix},\ C_T' = \begin{bmatrix} C_{ac}' \\ C_c' \end{bmatrix}. \tag{35}$$

△

The arguments used to prove the results above extend immediately to the case in which the probability density function a given random variable is defined on the whole real axis and its causal and anticausal parts are realized by a system in explicit form. The key point is that one has to consider a *signed sum* of the moments of the systems which realize the causal part and the anticausal part of the probability density function of the random variable of interest. For brevity we do not repeat other versions of these results for probability density functions realized by systems in explicit form; instead, we consider the following important example.

Example 6 (The normal distribution). The probability density function of a random variable X having a normal distribution with parameter $\sigma \in \mathbb{R}_{>0}$ is defined as

$$f_X : \mathbb{R} \to \mathbb{R},\ t \mapsto \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{t^2}{2\sigma^2}}. \tag{36}$$

The causal part of f_X is $f_c : \mathbb{R} \to \mathbb{R}$, $t \mapsto \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{t^2}{2\sigma^2}}\delta_{-1}(t)$, while the anticausal part of f_X is $f_{ac} : \mathbb{R} \to \mathbb{R}$, $t \mapsto \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{t^2}{2\sigma^2}}\mathfrak{d}_{-1}(t)$. The causal part and the time reversal of the anti-causal part of f_X are both realized by the linear, time-varying system

$$\dot{x} = -\frac{t}{\sigma^2} x + \frac{1}{\sqrt{2\pi\sigma^2}} u, \quad y = x, \tag{37}$$

in which $x(t) \in \mathbb{R}$, $u(t) \in \mathbb{R}$, $y(t) \in \mathbb{R}$. Consider now the interconnection of system (6) and of system (37), with $v = y$, set $s^\star = 0$ and note that Assumptions 1, 2, 3 and 4 hold. In analogy with Example 4, noting that equation (15) boils down to (25) gives $\Upsilon_\infty(t) = e^{St} H_+(t)$, with $H_+(t)$ defined as in (26). For a suitable signature matrix Ψ_k, defining $\Upsilon_{T,\infty}(t) = \Upsilon_\infty(t) + \Psi_k \Upsilon_\infty(t)$ and noting that $x(t) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{t^2}{2\sigma^2}}\delta_{-1}(t)$, by Definitions 3 and 4, allows to conclude (22) holds, with $\mu_0 = 1$ and

$$\mu_k = \begin{cases} \sigma^k\,(k-1)!!, & \text{if } k \text{ is even,} \\ 0, & \text{if } k \text{ is odd.} \end{cases}$$

Generalising the results of Theorem 6, this shows that the moments up to the order $k \in \mathbb{Z}_{\geq 0}$ of the random variable X uniquely specify the moment of system (24) at zero as $t \to \infty$. Note also that $\Upsilon_{T,\infty}$ can be written as $\Upsilon_{T,\infty} = (I + \Psi_k)\,\Upsilon_\infty$, which, in a broad sense, is in agreement with Corollary 5. ▲

D. Probability density functions with compact support

We now concentrate on probability density functions with compact support. To begin with a limitation of the characterisation of the moments of a random variables in terms of explicit systems is illustrated through a simple example.

Example 7 (The uniform distribution). Suppose we wish to find a realization of the probability density function of a random variable X having a uniform distribution with parameters $a, b \in \mathbb{R}_{>0}$, with $a < b$, defined as

$$f_X : \mathbb{R} \to \mathbb{R},\ t \mapsto \frac{1}{b-a} \mathbb{1}_{[a,b]}(t). \tag{38}$$

Clearly, any explicit realization of the form (12) necessarily violates Assumption 1, since f_X is zero everywhere except on a compact subset of the real line. As a result, the theory developed in Section III-B does not apply. However, the probability density function (38) can be also interpreted as the impulse response of the linear, time-delay system with discrete constant delays

$$\dot{x} = \frac{1}{b-a}\,(u_a - u_b), \quad y = x, \tag{39}$$

i.e. by system (8) with $x(t) \in \mathbb{R}$, $u(t) \in \mathbb{R}$, $y(t) \in \mathbb{R}$, $\tau_1 = a$, $\tau_2 = b$, $B_1 = \frac{1}{b-a}$, $B_2 = -\frac{1}{b-a}$ and $C_0 = 1$, $A(s) = 0$, $B(s) = \frac{1}{b-a}(e^{-sa} - e^{-sb})$, $C(s) = 1$. Note that the moments of system (39) are not classically defined at zero, since $0 \in \sigma(A(s))$. However, since zero is a removable singularity[9] of the transfer function of system (39), the moments of system (39) can be defined and characterised by means of Sylvester equations and impulse responses using the notions and results introduced in [85 – 87]. In particular, the moments of system (39) at zero satisfy the identity

$$[\,\eta_k(0)\ \cdots\ \eta_1(0)\ \eta_0(0)\,]' = \Psi_k \left(\frac{e^{-J_0 a} - e^{-J_0 b}}{b-a}\right) \Upsilon_k, \quad (40)$$

for every $k \in \mathbb{Z}_{\geq 0}$, with $\Psi_k \in \mathbb{R}^{(k+1)\times(k+1)}$ a signature matrix and $\Upsilon_k \in \mathbb{R}^{k+1}$ a solution of the (Sylvester) equation

$$J_0 \Upsilon_k + e_{k+1} = 0, \quad (41)$$

To see this, note that

$$\eta_k(0) = \frac{a^{k+1} - b^{k+1}}{(k+1)!(b-a)}. \quad (42)$$

Exploiting the definition of matrix exponential, the identity (41) and the property

$$J_0^{j-1} e_{k+1} = \begin{cases} e_j, & \text{for } 1 \leq j \leq k+1, \\ 0, & \text{for } j \geq k+2, \end{cases}$$

allows to conclude

$$\begin{aligned} \left(\frac{e^{-J_0 a} - e^{-J_0 b}}{b-a}\right) \Upsilon_k &= \sum_{j=0}^{\infty} \frac{(-1)^j}{j!} \frac{a^j - b^j}{b-a} J_0^j \Upsilon_k \\ &= \sum_{j=1}^{k+1} \frac{(-1)^{j+1}}{j!} \frac{a^j - b^j}{b-a} J_0^{j-1} e_{k+1} \\ &= \sum_{j=1}^{k+1} \frac{(-1)^{j+1}}{j!} \frac{a^j - b^j}{b-a} e_j \end{aligned}$$

which, in view of (42), proves the identity (40). We emphasise that, in line with the results developed for probability density functions realized by linear systems, the relation (20) between the moments of f_X and the moments of the corresponding realization is satisfied: a one-to-one correspondence exists between the moments of the random variable X and the moments of system (39), since $\mu_k = \frac{a^{k+1} - b^{k+1}}{(k+1)(b-a)}$. ▲

The main reason why it is possible to characterise in systems-theoretic terms the moments of a random variable having a uniform distribution is that zero is a removable singularity of the transfer function of the associated time-delay system. This observation allows to generalise the argument used in Example 7 to treat random variables the probability density function of which has compact support and is polynomial on the complement of its zero set. To see this, consider a random variable X having a probability density function of the form

$$f_X : \mathbb{R} \to \mathbb{R},\ t \mapsto q(t) \mathbb{1}_{[a,b]}(t), \quad (43)$$

in which $a, b \in \mathbb{R}_{>0}$, with $a < b$, and $q(t) = \sum_{k=0}^{\nu-1} q_k t^k$, with $q_{\nu-1} \neq 0$. Defining $Q_1(s) = \sum_{k=0}^{\nu-1} q_1^k s^k = \sum_{k=0}^{\nu-1} \sum_{i=0}^{k} \frac{k!}{i!} a^i s^{\nu-k+i-1}$ and $Q_2(s) = \sum_{k=0}^{\nu-1} q_2^k s^k = \sum_{k=0}^{\nu-1} \sum_{i=0}^{k} \frac{k!}{i!} b^i s^{\nu-k+i-1}$ the Laplace transform of (43) can be written as $\mathcal{L}\{f_X\} = \frac{Q_1(s)e^{-sa} - Q_2(s)e^{-sb}}{s^\nu}$ and has a removable singularity at zero if and only if $\sum_{k=0}^{l} (-1)^k \left(q_1^k a^{l-k} - q_2^k b^{l-k}\right) = 0$ for all $l \in \{0, 1, \ldots, \nu-1\}$. Under these conditions, the probability density function (43) can be realised by system (8) setting, *e.g.*, $\tau_1 = a$, $\tau_2 = b$,

$$A_0 = \begin{bmatrix} J_0 & 0 \\ 0 & J_0 \end{bmatrix},\ B_1 = \begin{bmatrix} e_\nu \\ 0 \end{bmatrix},\ B_2 = \begin{bmatrix} 0 \\ e_\nu \end{bmatrix},$$

$$C_1 = \begin{bmatrix} q_1^0 & q_1^1 & \cdots & q_1^{\nu-1} & 0 & 0 & \cdots & 0 \end{bmatrix},$$

$$C_2 = \begin{bmatrix} 0 & 0 & \cdots & 0 & q_2^0 & q_2^1 & \cdots & q_2^{\nu-1} \end{bmatrix},$$

and the moments of the random variable X can be shown to be in one-to-one correspondence with the moments at zero of such system. To illustrate this point, we consider a simple example.

Example 8 (The triangular distribution). The probability density function of a random variable X having a triangular distribution with parameter $\tau \in \mathbb{R}_{>0}$ is defined as

$$f_X : \mathbb{R} \to \mathbb{R},\ t \mapsto \frac{1}{\tau} \max\left(0, 1 - \frac{|t|}{\tau}\right). \quad (44)$$

The causal part of f_X is $f_c : \mathbb{R} \to \mathbb{R}$, $t \mapsto \frac{1}{\tau}\left(1 - \frac{t}{\tau}\right) \mathbb{1}_{[0,\tau]}(t)$, while the anticausal part of f_X is $f_{ac} : \mathbb{R} \to \mathbb{R}$, $t \mapsto \frac{1}{\tau}\left(1 + \frac{t}{\tau}\right) \mathbb{1}_{[-\tau,0]}(t)$. The causal part and the time reversal of the anti-causal part of f_X are both realized by the time-delay system

$$\dot{x} = e_2 u + e_4 u_\tau, \quad y = \frac{\tau e_2' - e_1'}{\tau^2} x + \frac{e_3'}{\tau^2} x_\tau, \quad (45)$$

i.e. by system (8) with $x(t) \in \mathbb{R}^4$, $u(t) \in \mathbb{R}$, $y(t) \in \mathbb{R}$, $\tau_1 = 0$, $\tau_2 = \tau$, $A_0 = 0$, $B_1 = e_2$, $B_2 = e_4$, $C_1 = \frac{\tau e_2' - e_1'}{\tau^2}$, $C_2 = \frac{e_3'}{\tau^2}$. The moment at zero of order $k \in \mathbb{Z}_{\geq 0}$ of the system is

$$\eta_k(0) = \frac{(-1)^k \tau^k + \tau^k}{2(k+2)!}, \quad (46)$$

which is consistent with the identity (20), since the moment of order $k \in \mathbb{Z}_{\geq 0}$ of the random variable X reads as

$$\mu_k = \frac{(-1)^k \tau^k + \tau^k}{2(k+2)(k+1)}. \quad (47)$$

This also implies that a one-to-one correspondence exists between the moments of the random variable X and the moments of the system. We emphasise that exploiting the argument of Example 7 the moments of the system (and thus those of the random variable X) may be computed using the formula (11). ▲

[9] See, *e.g.*, [84].

IV. Applications

This section contains a series of applications of the proposed ideas. We first focus on the identifiability of probability density functions admitting a linear realization. Then, a systems-theoretic interpretation for sums of independent random variables, the notion of mixture distribution and basic results from renewal theory are provided. Finally, connections between the approximation of probability density functions and the model reduction problem are studied.

A. Identifiability of probability density functions with linear realizations

We begin this section considering the case in which the probability density function of a given random variable is parameterized by a fixed set of parameters. In other words, while in the previous sections the parameters of probability density functions have been assumed to be known, in this section parameters are constant *unknown* quantities, which in principle can (or must) be estimated. In particular, we study the identifiability of parametric families of probability density functions the elements of which admit a linear realization. This is important, for example, in the context of parametric estimation, where identifiability allows to avoid redundant parametrisations and to achieve consistency of estimates [50, 51, 88].

Let Θ be an open subset of $\mathbb{R}^d$ representing the parameter set and let $\mathcal{F}_X$ be a family of probability density functions defined on the real axis and associated with a random variable X. Every element $f_X \in \mathcal{F}_X$ is a probability density function $t \mapsto f_X(t;\theta)$ which is known once the element $\theta \in \Theta$ has been fixed.

Definition 7. [88] The parameters $\theta_1 \in \Theta$ and $\theta_2 \in \Theta$ are observationally equivalent if $f_X(t;\theta_1) = f_X(t;\theta_2)$ for almost all[10] $t \in \mathbb{R}$.

The notion of observational equivalence induces an equivalence relation on the parameter set, defined as $\theta_1 \sim \theta_2$ if $\theta_1 \in \Theta$ and $\theta_2 \in \Theta$ are observationally equivalent. The parameter set is therefore partitioned into equivalence classes the cardinality of which determines the identifiability of the family of probability density functions considered, as specified by the following definition.

Definition 8. The family of probability density functions $\mathcal{F}_X$ is identifiable if $\theta_1 \sim \theta_2$ implies $\theta_1 = \theta_2$ for all $\theta_1 \in \Theta$ and $\theta_2 \in \Theta$.

A characterisation of the identifiability of a family of probability density functions admitting a linear realization can be given by means of the systems-theoretic notion of minimality. To this end, note that the description of probability density functions as impulse responses has an inherent non-uniqueness issue, since algebraically equivalent[11] linear systems have the same impulse response. However, a one-to-one correspondence between impulse responses and their realizations can be established resorting to canonical forms [80], such as the observer canonical form, defined by constant matrices $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^n$ and $C \in \mathbb{R}^{1\times n}$ of the form

$$A = \begin{bmatrix} -\alpha_n & 1 & 0 & \dots & 0 \\ \vdots & 0 & 1 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ -\alpha_2 & 0 & \dots & 0 & 1 \\ -\alpha_1 & 0 & \dots & 0 & 0 \end{bmatrix}, \; B = \begin{bmatrix} \beta_n \\ \vdots \\ \vdots \\ \beta_2 \\ \beta_1 \end{bmatrix}, \; C = e_1', \qquad (48)$$

with $\alpha = (\alpha_1, \dots, \alpha_n) \in \mathbb{R}^n$ and $\beta = (\beta_1, \dots, \beta_n) \in \mathbb{R}^n$.

With these premises, we may recover the following well-known result [51].

Lemma 6. Let Θ be an open subset of $\mathbb{R}^d$, representing the parameter set, and let $\mathcal{F}_X$ be a family of probability density functions defined on the real axis and associated with a random variable X. Assume every $f_X \in \mathcal{F}_X$ is realized by system (1) with $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^n$ and $C \in \mathbb{R}^{1\times n}$ as in (48) and let $\theta = (\alpha, \beta) \in \Theta$. Then the family of probability density functions $\mathcal{F}_X$ is identifiable if and only if every pair (A, B) is reachable.

Proof. Note that the family of probability density functions $\mathcal{F}_X$ is identifiable if and only if for every $f_X \in \mathcal{F}_X$ the map

$$(\alpha, \beta) \mapsto F_X(s;\alpha,\beta) = \frac{\beta_n s^{n-1} + \dots + \beta_2 s + \beta_1}{s^n + \alpha_n s^{n-1} + \dots + \alpha_2 s + \alpha_1}$$

with $F_X(s;\alpha,\beta) = \mathcal{L}\{f_X\}$, is injective. This, in turn, corresponds to the numerator and denominator of the rational function $F_X(s;\alpha,\beta)$ being coprime. As a result, the identifiability of the family $\mathcal{F}_X$ is equivalent to the minimality of system (1) and, by observability of the pair (A, C), to the reachability of the pair (A, B), which proves the claim. □

Remark 7. A dual result can be proved using the controllability canonical form as long as observability, and hence minimality, is enforced. This suggests that the identifiability of a family of probability density functions admitting a linear realization is equivalent to the minimality of a given canonical realization, which can be thus taken as the *definition* of identifiability. △

B. Sums of independent random variables

A classical theorem of probability theory states that the probability density function of the sum of two jointly continuous, independent random variables is given by the convolution of their probability density functions (see, *e.g.*, [73]). This result can be given a simple systems-theoretic interpretation.

Theorem 7. Let X_1 and X_2 be jointly continuous, independent random variables with probability density functions f_{X_1} and f_{X_2} realized by the minimal system

$$\dot{x}_1 = A_1 x_1 + B_1 u_1, \quad y_1 = C_1 x_1, \qquad (49)$$

and the minimal system

$$\dot{x}_2 = A_2 x_2 + B_2 u_2, \quad y_2 = C_2 x_2, \qquad (50)$$

[10] A property is satisfied for almost all $t \in \mathbb{R}$ if the set where the property does not hold has Lebesgue measure equal to zero.

[11] The single-input, single-output, continuous-time, linear, time-invariant systems $\dot{x}_1 = A_1 x_1 + B_1 u_1$, $y_1 = C_1 x_1$, with $x_1(t) \in \mathbb{R}^n$, $u_1(t) \in \mathbb{R}$, $y_1(t) \in \mathbb{R}$ and $\dot{x}_2 = A_2 x_2 + B_2 u_2$, $y_2 = C_2 x_2$, with $x_2(t) \in \mathbb{R}^n$, $u_2(t) \in \mathbb{R}$, $y_2(t) \in \mathbb{R}$ are algebraically equivalent if there exists a non-singular matrix $T \in \mathbb{R}^{n\times n}$ such that $A_2 = TA_1T^{-1}$, $B_2 = TB_1$, $C_2 = C_1T^{-1}$.

with $x_j(t) \in \mathbb{R}^{n_j}$, $u_j(t) \in \mathbb{R}$, $y_j(t) \in \mathbb{R}$, and $A_j \in \mathbb{R}^{n_j \times n_j}$, $B_j \in \mathbb{R}^{n_j \times 1}$ and $C_j \in \mathbb{R}^{1 \times n_j}$ constant matrices for $j \in \{1, 2\}$, respectively. Then the probability density functions of the random variable $Y = X_1 + X_2$ is realized by the interconnection of system (49) and system (50) with $u_1 = y_2$.

Proof. Recall that the probability density function of the sum of two jointly continuous, independent random variables is given by the convolution of their probability density functions [73, Theorem 6.38], *i.e.* $f_Y = f_{X_1} * f_{X_2}$. Taking the Laplace transform on both sides, this implies $\mathcal{L}\{f_Y\} = \mathcal{L}\{f_{X_1}\}\mathcal{L}\{f_{X_1}\}$. Thus, the probability density function f_Y is realized by the interconnection of systems (49) and (50) with $u_1 = y_2$, since the transfer function associated with the probability density function f_Y is the product of the transfer functions associated with the probability density functions f_{X_1} and f_{X_2}. □

The following result is an immediate extension of Theorem 7.

Corollary 6. Let $X_1, X_2, \ldots, X_N$ be jointly continuous, independent random variables. Assume the probability density function f_{X_j} is realized by the minimal system

$$\dot{x}_j = A_j x_j + B_j u_j, \quad y_j = C_j x_j, \tag{51}$$

with $x_j(t) \in \mathbb{R}^{n_j}$, $u_j(t) \in \mathbb{R}$, $y_j(t) \in \mathbb{R}$, and $A_j \in \mathbb{R}^{n_j \times n_j}$, $B_j \in \mathbb{R}^{n_j \times 1}$ and $C_j \in \mathbb{R}^{1 \times n_j}$ constant matrices for $j \in \{1, 2, \ldots, N\}$, respectively. Then the probability density functions of the random variable $Y = X_1 + X_2 + \ldots + X_N$ is realized by the interconnection of the family of systems (51) with $u_1 = y_2$, $u_2 = y_3$, $\ldots$, $u_{N-1} = y_N$.

Example 9 (The Erlang distribution). Suppose we wish to show that the probability density function f_Y of the random variable $Y = X_1 + X_2 + \ldots + X_N$, in which $N \in \mathbb{Z}_{>0}$ and $X_1, X_2, \ldots, X_N$ are jointly continuous, independent random variables having exponential distribution with parameter $\lambda \in \mathbb{R}_{>0}$, is that of an Erlang distribution with parameters λ and N, defined as

$$f_Y : \mathbb{R} \to \mathbb{R} : t \mapsto \frac{\lambda^N}{(N-1)!} t^{N-1} e^{-\lambda t} \delta_{-1}(t). \tag{52}$$

Recall that a minimal realization of the probability density function f_{X_j} of the random variable X_j is described by system (21) for all $j \in \{1, 2, \ldots, N\}$. Thus, by Corollary 6, a realization of the probability density function f_Y is given by system (1), in which

$$A = J_{-\lambda}, \quad B = e_N, \quad C = \frac{\lambda^N}{(N-1)!} e_1'. \tag{53}$$

We conclude this example noting that a direct computation shows that the probability density function f_Y is indeed given by (52). ▲

Remark 8. A random variable Y is decomposable if there exist $N \in \mathbb{Z}_{>0}$, with $N \geq 2$, and jointly continuous, independent random variables $X_1, X_2, \ldots, X_N$ such that $Y = X_1 + X_2 + \ldots + X_N$ [89]. In case the random variables $X_1, X_2, \ldots, X_N$ are also identically distributed, then Y is said to be divisible [89]. The notions of decomposability and of divisibility play an important role in probability theory [90], particularly in the analysis of Lévy processes [7, 67]. In light of Theorem 7 of Corollary 6, these notions can be characterised in systems-theoretic terms. In particular, decomposability (and divisibility) of a random variable are related to the possibility of describing the corresponding system as the series interconnection of finitely many (and possibly identical) systems. △

The following result provides a systems-theoretic necessary and sufficient condition which ensures the identifiability of a family of probability density function the elements of which can be represented as the sum of random variables with probability density functions admitting a linear realization.

Theorem 8. Let $\mathcal{F}_X$ be a family of probability density functions the elements of which can be realized as the sum of the probability density functions f_{X_1} and f_{X_2}. Assume f_{X_1} and f_{X_2} are realized by the minimal systems (49) and (50), respectively. Then the family of probability density functions $\mathcal{F}_X$ is identifiable if and only if

(i) the polynomials $C_1 \operatorname{adj}(sI - A_1)B_1$ and $\det(sI - A_2)$ have no common roots and

(ii) the polynomials $C_2 \operatorname{adj}(sI - A_2)B_2$ and $\det(sI - A_1)$ have no common roots.

Proof. The identifiability of the family of probability density functions $\mathcal{F}_X$ is equivalent to the minimality of the series interconnection of the minimal systems (49) and (50), which, in turn, is equivalent to conditions (i) and (ii) [80, 91]. □

C. Mixture distributions

We have seen that the sum of two jointly continuous independent random variables has a natural interpretation in terms of the series interconnection of the realizations of their probability density functions. To provide a probabilistic counterpart of the notion of parallel interconnection we recall the following definition, which is adapted from [92].

Definition 9. A random variable Z with probability density function f_Z is said to arise from a finite mixture distribution if there exist $N \in \mathbb{Z}_{>0}$ and jointly continuous random variables $X_1, X_2, \ldots, X_N$ with probability density functions $f_{X_1}, f_{X_2}, \ldots, f_{X_N}$ such that the probability density function f_Z satisfies $f_Z = w_1 f_{X_1} + w_2 f_{X_2} + \cdots + w_N f_{X_N}$, for some $w = (w_1, w_2, \ldots, w_n) \in \Delta^{n-1}$. N is referred to as the number of components of f_Z, $f_{X_1}, f_{X_2}, \ldots, f_{X_N}$ are referred to as the components of f_Z, and $w_1, w_2, \ldots, w_N$ are referred to as the weights of f_Z.

Theorem 9. Under the hypotheses of Theorem 7, if the random variable Z with probability density function f_Z arises from a finite mixture distribution with components f_{X_1} and f_{X_2} and weights $w_1 \neq 0$ and $w_2 \neq 0$, then the probability density function f_Z is realized by the interconnection of system (49) and system (50) with $u = u_1 = u_2$ and $y = w_1 y_1 + w_2 y_2$.

Proof. By hypothesis, the probability density function f_Z arises from a finite mixture distribution with components f_{X_1} and f_{X_2} and weights w_1 and w_2, *i.e.* $f_Z = w_1 f_{X_1} + w_2 f_{X_2}$. Taking the Laplace transform on both sides, by linearity, yields $\mathcal{L}\{f_Z\} = w_1 \mathcal{L}\{f_{X_1}\} + w_2 \mathcal{L}\{f_{X_2}\}$. Thus, the probability density function f_Z is realized by the interconnection of system (49) and

system (50) with $u = u_1 = u_2$ and $y = w_1 y_1 + w_2 y_2$, as desired. □

Example 10 ($G/H_2/1$ queueing system). Consider the $G/H_2/1$ queueing system in Fig. 1, in which the arrival process is a general random process and the service process is governed by a two-phase hyper-exponential random variable X. A customer accesses either the service offered by the first server at rate $\lambda_1 \in \mathbb{R}_{>0}$ with probability $p_1 \in (0,1)$ or the service offered by the second server at rate $\lambda_2 \in \mathbb{R}_{>0}$ with probability $p_2 = 1 - p_1$. The probability density function of the random variable X which represents the service time, *i.e.* the time spent by an arbitrary customer in the service process, is $f_X(t) = p_1 \lambda_1 e^{-\lambda_1 t} \delta_{-1}(t) + p_2 \lambda_2 e^{-\lambda_2 t} \delta_{-1}(t)$ for all $t \in \mathbb{R}_{\geq 0}$. In view of Theorem 9, the probability density function of the service time is realized by system (1), with

$$A = \begin{bmatrix} -\lambda_1 & 0 \\ 0 & -\lambda_2 \end{bmatrix}, \; B = \begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix}, \; C = \begin{bmatrix} p_1 \\ p_2 \end{bmatrix}'. \tag{54}$$

▲

A straightforward generalisation of Theorem 9 is given by the following result.

Corollary 7. Under the hypotheses of Corollary 6, let the random variable Z with probability density function f_Z arise from a finite mixture distribution with components $f_{X_1}, f_{X_2}, \ldots, f_{X_N}$ and weights $w_1, w_2, \ldots, w_N \neq 0$. Then the probability density function f_Z is realized by the interconnection of systems (51) with $u = u_1 = u_2 = \ldots = u_N$ and $y = w_1 y_1 + w_2 y_2 + \ldots + w_N y_N$.

We conclude this section presenting a systems-theoretic necessary and sufficient condition which guarantees the identifiability of finite mixtures admitting a linear realization (see also [92]).

Theorem 10. Let $\mathcal{F}_X$ be a family of probability density functions, the elements of which arise from a finite mixture distribution with components f_{X_1} and f_{X_2} and weights $w_1 \in \mathbb{R}_{>0}$ and $w_2 \in \mathbb{R}_{>0}$. Assume f_{X_1} and f_{X_2} are realized by the minimal systems (49) and (50), respectively. Then the family of probability density functions $\mathcal{F}_X$ is identifiable if and only if $\sigma(A_1) \cup \sigma(A_2) = \emptyset$.

Proof. The family of probability density functions $\mathcal{F}_X$ is equivalent to the minimality of the parallel interconnection with weights $w_1 \in \mathbb{R}_{>0}$ and $w_2 \in \mathbb{R}_{>0}$ of the minimal systems (49) and (50), which, in turn, is equivalent to $\sigma(A_1) \cup \sigma(A_2) = \emptyset$ [80, 91]. □

Example 11 ($G/H_2/1$ queueing system, continued). Suppose we are interested in finding conditions under which the family of probability density functions

$$\mathcal{F}_X = \left\{ t \mapsto f_X(t;\lambda) = \sum_{j=1}^{2} p_j \lambda_j e^{-\lambda_j t} \delta_{-1}(t), \; \lambda \in \mathbb{R}^2_{>0} \right\},$$

which describes the service time of the $G/H_2/1$ queueing system displayed in Fig. 1, is identifiable. Note that the probability density function of the service time is realized by system (1), with matrices defined as in (54). By Theorem 10, since $p_1, p_2 \in (0,1)$, the family of probability density functions $\mathcal{F}_X$ is identifiable if and only if $\lambda \in \mathbb{R}^2$ does not lie on the bisector of the first quadrant, *i.e.* $\lambda_1 \neq \lambda_2$. Note that in the case $\lambda_1 = \lambda_2 = \bar{\lambda} \in \mathbb{R}_{>0}$ a customer accesses the service offered by a server at rate $\bar{\lambda}$ with probability one and hence the model is overparameterised. In other words, the queueing system in question may be equivalently described by the $G/M/1$ queueing system displayed in Fig. 2, in which the service process is governed by an exponential random variable X with parameter $\bar{\lambda}$. As anticipated in Remark 7, this phenomenon is also captured by the systems-theoretic notion of minimality: in accordance with Theorem 10 for $\lambda_1 = \lambda_2$ the system (54) is not minimal. ▲

D. Renewal processes

We complete this section showing that elementary results from renewal theory [6, 7, 69] can be translated in the language of systems theory using the notion of feedback interconnection.

Definition 10. [7] A sequence of random variables $\{S_j\}_{j \in \mathbb{Z}_{\geq 0}}$ constitutes a renewal process if it is of the form[12] $S_j = T_1 + T_2 + \ldots + T_j$, where $\{T_j\}_{j \in \mathbb{Z}_{>0}}$ is a sequence of mutually independent random variables with a common distribution F such that $F(0) = 0$.

The random variable S_j in the above definition is often referred to as (the j-th) renewal, while the elements of the sequence $\{T_j\}_{j \in \mathbb{Z}_{>0}}$ are referred to as waiting times [7, 69]. The common distribution of the waiting times of a renewal process is called the waiting-time distribution [69].

The probabilistic behaviour of a renewal process $\{S_j\}_{j \in \mathbb{Z}_{\geq 0}}$ is closely related to the random variable N_t, with $t \in \mathbb{R}_{>0}$, defined as the largest $j \in \mathbb{Z}_{\geq 0}$ for which $S_j \leq t$ [93]. The random variable N_t describes the number of renewals occurred by time t and its expected value $H(t) = \mathbb{E}[N_t]$, referred to as the renewal function, satisfies the integral equation of renewal theory [93]. Moreover, if the waiting-time distribution of the renewal process is absolutely continuous then the renewal density, defined as $h(t) = \dot{H}(t)$ for all $t \in \mathbb{R}_{>0}$, satisfies the renewal density integral equation [93], *i.e.*

$$h(t) = f(t) + \int_0^t h(t - \zeta) f(\zeta) d\zeta, \tag{55}$$

in which f is the derivative of the waiting-time distribution.

Theorem 11. Let $\{S_j\}_{j \in \mathbb{Z}_{\geq 0}}$ be a renewal process. Assume the waiting-time distribution of the renewal process $\{S_j\}_{j \in \mathbb{Z}_{\geq 0}}$ admits a probability density function f which is realized by the minimal system (1). Then the renewal density h of the renewal process $\{S_j\}_{j \in \mathbb{Z}_{\geq 0}}$ is realized by the system obtained from system (1) with input $v(t) \in \mathbb{R}$, output $y(t) \in \mathbb{R}$, and interconnection equation $u = v + y$.

Proof. To begin with note that under the stated assumptions the renewal density h of the renewal process $\{S_j\}_{j \in \mathbb{Z}_{\geq 0}}$ is well-defined [69, Proposition 2.7]. In addition, the renewal density h satisfies the renewal density integral equation (55). This implies that the Laplace transform of the renewal density

[12] By convention, $S_0 = 0$ and 0 counts as zero occurrences [7].

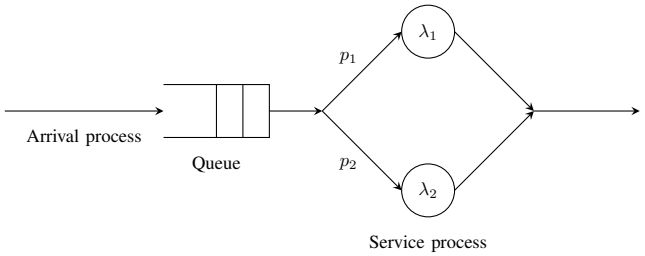


Fig. 1: $G/H_2/1$ queueing system.

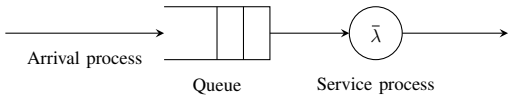


Fig. 2: $G/M/1$ queueing system.

is such that $\mathcal{L}\{h\} = \mathcal{L}\{f\}/(1-\mathcal{L}\{f\})$ [93, p.252]. Thus, since by hypothesis $\mathcal{L}\{f\}$ coincides with the transfer function of system (1), the renewal density h is realized by the system obtained from system (1) with input $v(t) \in \mathbb{R}$, output $y(t) \in \mathbb{R}$, and interconnection equation $u = v + y$. □

Example 12 (Poisson processes). Poisson processes are renewal processes in which the waiting times have an exponential distribution [70]. In other words, a renewal process is a Poisson process if there exists $\lambda \in \mathbb{R}_{>0}$ such that the probability density function of each waiting time is $f(t) = \lambda e^{-\lambda t}\delta_{-1}(t)$, for all $t \in \mathbb{R}_{\geq 0}$. In view of Theorem 11, the renewal density of the process is realized by the system obtained from system (21) with input $v(t) \in \mathbb{R}$, output $y(t) \in \mathbb{R}$, and interconnection equation $u = v + y$, *i.e.* by system (1), with $A = 0$, $B = \lambda$, $C = 1$. Note that the impulse response of the system is given by $h(t) = \lambda\delta_{-1}(t)$ for all $t \in \mathbb{R}$, which is consistent with the fact that the renewal function of a Poisson process can be written as $H(t) = \lambda t$ for all $t \in \mathbb{R}_{\geq 0}$ [94]. ▲

E. On the approximation of probability density functions

This section investigates some connections between the approximation of probability density functions and the model reduction problem [25]. In particular we show that these problems are essentially the same problem when probability density functions are regarded as impulse responses. As a guiding example, we consider phase-type distributions [30, 31, 33, 34], which play an important role in the analysis of queuing networks [94] and can be represented by a random variable describing the time until absorption of a Markov chain with one absorbing state [95]. Note, however, that similar considerations can be performed for more general classes of probability density functions.

Consider a continuous-time Markov chain over the set $\mathcal{S} = \{0, 1, \ldots, n\}$, with $n \in \mathbb{Z}_{>0}$, in which 0 is an absorbing state and $1, \ldots, n$ are transient states. The random variable X which characterises the time until absorption is described by a continuous phase-type distribution represented by the Q-matrix [37]

$$Q = \begin{bmatrix} 0 & 0 \\ \overline{Q}_0 & \overline{Q} \end{bmatrix},$$

in which $S \in \mathbb{R}^{n\times n}$ is such that $\overline{Q}_0 = -\overline{Q}\mathbb{1}$. Assuming that the initial probability of the Markov chain is

$$\pi_0 = \begin{bmatrix} 0 & \alpha \end{bmatrix},$$

with $\alpha \in \mathbb{R}^{1\times n}$ such that $\alpha\mathbb{1} = 1$, the probability density function of X reads as

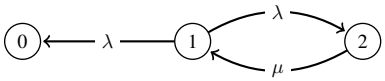
$$f_X(t) = \overline{Q}_0' e^{\overline{Q}' t}\alpha'\delta_{-1}(t), \tag{56}$$

for every $t \in \mathbb{R}$. Note that (56) can be regarded as the impulse response of system (1), with $A = \overline{Q}'$, $B = \alpha'$ and $C = \overline{Q}_0'$. This indicates that the problem of approximating the probability density function (56) can be regarded as the problem of approximating the impulse response of system (1) or, equivalently, the problem of constructing a reduced order model of system (1) [25]. In particular, approximating the probability density function (56) by another phase-type distribution boils down to constructing a system

$$\dot{\xi} = F\xi + Gv, \quad \psi = H\xi, \tag{57}$$

in which $\xi(t) \in \mathbb{R}^\nu$, with $\nu < n$, $v(t) \in \mathbb{R}$, $\psi(t) \in \mathbb{R}$ and $F \in \mathbb{R}^{\nu\times\nu}$, $G \in \mathbb{R}^{\nu\times 1}$ and $H \in \mathbb{R}^{1\times\nu}$ are constant matrices such that $\mathbb{1}'G = 1$ and $H = -\mathbb{1}'F$. To illustrate this point we consider a simple example and exploit the model reduction technique devised in [61, 62] to obtain reduced order models (*i.e.* approximations of a given probability density function) which match a prescribed number of (systems-theoretic and, thus, probabilistic) moments. Note, however, that different model reduction techniques can be used to solve the problem in question (see [25] and references therein for an overview of available model reduction techniques).

Example 13. Consider the Markov chain described by the diagram



The Q-matrix of the Markov chain is

$$Q = \begin{bmatrix} 0 & 0 & 0 \\ \lambda & -\lambda & 0 \\ \mu & \lambda & -(\lambda+\mu) \end{bmatrix},$$

and

$$\overline{Q}_0 = \begin{bmatrix} \lambda \\ \mu \end{bmatrix}, \qquad \overline{Q} = \begin{bmatrix} -\lambda & 0 \\ \lambda & -(\lambda+\mu) \end{bmatrix}.$$

Let $\alpha = [\,0\ 0\ 1\,]$ be the the initial condition of the Markov chain and let X be the random variable which characterises the time until absorption. A realization of the probability density function of the random variable X is given by system (1), with

$$A = \begin{bmatrix} -\lambda & \lambda \\ 0 & -(\lambda+\mu) \end{bmatrix}, \quad B = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad C = \begin{bmatrix} \lambda & \mu \end{bmatrix}.$$

TABLE I: Random variables and associated probability density function, parameters and realization.

Exponential	$\lambda e^{-\lambda t}\delta_{-1}(t)$	$\lambda \in \mathbb{R}_{>0}$	(21)
Half-normal	$\sqrt{\frac{2}{\pi\sigma^2}}e^{-\frac{t^2}{2\sigma^2}}\delta_{-1}(t)$	$\sigma \in \mathbb{R}_{>0}$	(24)
Laplace	$\frac{\lambda}{2}e^{-\lambda\lvert t\rvert}$	$\lambda \in \mathbb{R}_{>0}$	(32)
Normal	$\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{t^2}{2\sigma^2}}$	$\sigma \in \mathbb{R}_{>0}$	(37)
Uniform	$\frac{1}{b-a}\mathbb{1}_{[a,b]}(t)$	$a, b \in \mathbb{R}$ $(a < b)$	(39)
Triangular	$\frac{1}{\tau}\max\left(0, 1-\frac{\lvert t\rvert}{\tau}\right)$	$\tau \in \mathbb{R}_{>0}$	(45)
Erlang	$\frac{\lambda^n t^{n-1}e^{-\lambda t}}{(n-1)!}\delta_{-1}(t)$	$\lambda \in \mathbb{R}_{>0}$	(53)

Following [62], to construct a reduced order model which matches the moment of order one of the system at zero one needs to solve the Sylvester equation (4), with $S = 0$ and $M = 1$, which gives $\Upsilon_1 = \begin{bmatrix} -1 & -1 \end{bmatrix}$. Then one defines reduced order model (57) as

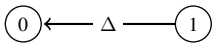
$$F = S - M\Delta, \quad G = -\Upsilon_1 B, \quad H = \Delta,$$

with $\Delta \in (0, 1)$ a free parameter that can be assigned, yielding

$$F = -\Delta, \qquad G = 1, \qquad H = \Delta.$$

Note that the structure of the original system is preserved in the reduced order model, since $\mathbb{1}'G = 1$ and $H = -\mathbb{1}'F$. Moreover, in agreement with the results of [62], the moment at zero of the reduced order model coincides with the moment at zero of the original system, regardless of the value of Δ.

From a probabilistic point of view, the impulse response of the reduced order model corresponds to the probability density function of the random variable $\widetilde{X}$ which quantifies the time until absorption of the "reduced" Markov chain described by the diagram



and the Q-matrix

$$\widetilde{Q} = \begin{bmatrix} 0 & 0 \\ \Delta & -\Delta \end{bmatrix}.$$

It is interesting to note that the "reduced" Markov chain can be interpreted as Markov chain built from the original Markov chain by aggregation of the states 2 and 3, thus showing the connection between the model reduction problem and the use of the concept of aggregation to reduce the state space of a Markov chain [96 – 98].

We conclude this example by emphasising that there is no natural choice of the parameter Δ. For example, one may select $\Delta = \lambda$ to ensure that in the "reduced" Markov chain the transition rate from 1 to 0 matches that of the original Markov chain. Another sensible choice is $\Delta = \frac{\lambda+\mu}{2}$ which guarantees that the moments of order two at zero of the reduced order model and of the original system coincide, so that the moments of order two of the random variables X and $\widetilde{X}$ also coincide. ▲

TABLE II: Correspondence between concepts of probability theory and of systems theory.

Sum of independent random variables	Series interconnection
Mixture distribution	Parallel interconnection
Renewal process	Feedback interconnection

V. Conclusion

Moments of continuous random variables admitting a probability density function have been studied. Under certain hypotheses, the moments of a random variable have been shown to be in one-to-one correspondence with the solutions of a Sylvester equation and with the steady-state output response of a specific interconnected system. The results established in this work have shown that, under certain assumptions, a system can be seen as an *alternative, equivalent* description of the probabilistic structure of a random variable. This, in turn, indicates that systems-theoretic techniques and tools can be used to revisit and shed new light on classical results of probability theory and statistics. Table I displays a short list of random variables along with their probability density functions, parameters and associated realization, while Table II summarises the correspondence between certain concepts of probability theory and their systems-theoretic counterpart.

The present work is a first step towards a unified understanding of the role of systems-theoretic concepts in probability theory and statistics. Several directions for interdisciplinary research are left open. For example, discrete-time systems [99] provide the right tool to carry out the analysis presented in this work for discrete random variables. Hybrid systems [100 , 101] may be used to deal with random variables the distribution function of which has both a discrete part and an absolutely continuous part. Moments of multivariate distributions may be studied resorting to systems described by PDEs [102 , 103], for continuous random variables, and nD systems [104], for discrete random variables. As a consequence, conditional probabilities may be characterised in systems-theoretic terms. Note, however, that modelling moments of multivariate distributions using PDEs might raise challenging computational issues. Further connections may be explored with positive systems [38] and, in particular, with Markov chains [5 , 6 , 8 , 35 – 37]. The interplay between the role of Hankel matrices in realization theory [79 – 81] and in probability theory [5 – 8] may be studied. The notions of information and of entropy may be investigated using the proposed framework. The discussion on the approximation of probability density functions which arise from phase-type distributions can be generalised to probability density functions with a systems-theoretic representation in explicit form, for which model reduction techniques are available [58 , 77]. Finally, a systems-theoretic counterpart of the method of moments [18] may be developed modifying existing data-driven model reduction methods [63].

References

[1] H. Cramér, *Mathematical methods of statistics*. Princeton, NJ: Princeton Univ. Press, 1946.

[2] M. G. Kendall and A. Stuart, *The advanced theory of statistics.* London, U.K.: Griffin, 1977, vol. I.

[3] R. J. Serfling, *Approximation theorems of mathematical statistics*. New York: Wiley, 1980.

[4] E. L. Lehmann and G. Casella, *Theory of point estimation (2nd edition)*. New York: Springer, 1998.

[5] J. L. Doob, *Stochastic processes*. New York: Wiley, 1953.

[6] W. Feller, *An introduction to probability theory and its applications (3rd edition)*. New York: Wiley, 1971, vol. 1.

[7] ——, *An introduction to probability theory and its applications (3rd edition)*. New York: Wiley, 1971, vol. 2.

[8] P. Billingsley, *Probability and measure (2nd edition)*. New York: Wiley, 1986.

[9] W. Rudin, *Principles of mathematical analysis*. New York: McGraw-Hill, 1964.

[10] T. J. Stieltjes, "Recherches sur les fractions continues," in *Ann. Fac. Sci. Toulouse Sci. Math. Sci. Phys.*, vol. 8, 1894, pp. 1–122.

[11] H. Hamburger, "Über eine erweiterung des stieltjesschen momenten-problems," *Math. Ann.*, vol. 81, pp. 235–319, 1920.

[12] F. Hausdorff, "Momentprobleme für ein endliches intervall," *Math. Z.*, vol. 16, pp. 220–248, 1923.

[13] Y. V. Vorobyev, *Methods of moments in applied mathematics (translated from russian by B. Seckler)*. New York: Gordon and Breach, 1965.

[14] H. Fischer, *A history of the central limit theorem: from classical to modern probability theory*. New York: Springer, 2010.

[15] M. Kac, "On some connection between probability theory and differential and integral equations," *Proc. 2nd Berkeley Sympos. Math. Stat. Prob*, 1951, pp. 189–215.

[16] E. P. Wigner, "On the distribution of the roots of certain symmetric matrices," *Ann. Math.*, vol. 67, pp. 325–327, 1958.

[17] S. Karlin and J. L. Mc Gregor, "The differential equations of birth-and-death processes, and the Stieltjes moment problem," *Trans. Amer. Math. Soc.*, vol. 85, no. 2, pp. 481–546, 1957.

[18] K. Pearson, "Contributions to the mathematical theory of evolution," *Phil. Trans. Roy. Soc.*, vol. 185a, pp. 71–110, 1894.

[19] J. H. Friedman and J. W. Tukey, "A projection pursuit algorithm for exploratory data analysis," *IEEE Trans. Comput.*, vol. 23, pp. 881–890, 1974.

[20] P. J. Huber, "Projection pursuit," *Ann. Stat.*, vol. 13, no. 2, pp. 435–475, 1985.

[21] J. H. Friedman, "Exploratory projection pursuit," *J. Amer. Stat. Ass.*, vol. 82, pp. 249–266, 1987.

[22] N. I. Akhiezer, *The classical moment problem*. New York: Hafner, 1965.

[23] S. Karlin and W. J. Studden, *Tchebycheff systems: with applications in analysis and statistics*. New York: Interscience, 1966.

[24] H. J. Landau, Ed., *Moments in mathematics*. Providence, RI: Amer. Math. Soc., 1987.

[25] A. C. Antoulas, *Approximation of large-scale dynamical systems*. Philadelphia, PA: SIAM, 2005.

[26] C. Commault and S. Mocanu, "Phase-type distributions and representations: some results and open problems for system theory," *Int. J. Control*, vol. 76, no. 6, pp. 566–580, 2003.

[27] L. Benvenuti and L. Farina, "A tutorial on the positive realization problem," *IEEE Trans. Autom. Control*, vol. 49, no. 5, pp. 651–664, 2004.

[28] B. Hanzon and R. J. Ober, "A state-space calculus for rational probability density functions and applications to non-gaussian filtering," *SIAM J. Control Optim.*, vol. 40, no. 3, pp. 724–740, 2001.

[29] ——, "State space calculations for discrete probability densities," *Linear Algebra Appl.*, vol. 350, no. 1, pp. 67–87, 2002.

[30] C. A. O'Cinneide, "Characterization of phase-type distributions," *Stoch. Models*, vol. 6, no. 1, pp. 1–57, 1990.

[31] ——, "Phase-type distributions: open problems and a few properties," *Stoch. Models*, vol. 15, no. 4, pp. 731–757, 1999.

[32] M. Bladt and M. F. Neuts, "Matrix-exponential distributions: Calculus and interpretations via flows," *Stoch. Models*, vol. 15, no. 4, pp. 731–757, 2003.

[33] M. F. Neuts, *Matrix-geometric solutions in stochastic models: an algorithmic approach*. Baltimore, MD: John Hopkins Univ. Press, 1981.

[34] G. Latouche and V. Ramaswami, *Introduction to matrix analytic methods in stochastic modeling*. Philadelphia, PA: SIAM, 1999.

[35] E. B. Dynkin, *Markov processes*. New York: Plenum, 1963.

[36] K. L. Chung, *Markov chains with stationary transition probabilities*. New York: Springer-Verlag, 1967.

[37] J. R. Norris, *Markov chains*. Cambridge, U.K.: Cambridge Univ. Press, 1998.

[38] L. Farina and S. Rinaldi, *Positive linear systems - theory and applications*. New York: Wiley, 2000.

[39] F. Gantmacher, *The theory of matrices*. New York: Chelsea, 1959, vol. II.

[40] E. Seneta, *Non-negative matrices and Markov chains (2nd edition)*. New York: Springer-Verlag, 1981.

[41] R. A. Horn and C. R. Johnson, *Matrix analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1985.

[42] H. Minc, *Nonnegative matrices*. New York: Wiley, 1988.

[43] C. G. Cassandras and S. Lafortune, *Introduction to discrete event systems*. New York: Kluwer, 1999.

[44] P. Brémaud, *Markov chains: Gibbs fields, Monte Carlo simulation, and queues*. New York: Springer-Verlag, 2001.

[45] M. Andretta, "Some considerations on the definition of risk based on concepts of systems theory and probability," *Risk Analysis*, vol. 34, no. 7, pp. 1184–1195, 2014.

[46] N. Leveson, "A systems approach to risk management through leading safety indicators," *Reliab. Eng. Syst. Saf.*, vol. 136, pp. 17–34, 2015.

[47] W. C. Elmore, "The transient response of damped linear networks with particular regard to wideband amplifiers," *J. App. Phys.*, vol. 19, no. 1, pp. 55–63, 1948.

[48] R. E. Kalman, "A new approach to linear filtering and prediction problems," *J. Basic Eng.*, vol. 82, pp. 35–45, 1960.

[49] R. E. Kalman and R. S. Bucy, "New results in linear filtering and prediction theory," *J. Basic Eng.*, vol. 83, pp. 95–108, 1961.

[50] T. D. Söderström and P. G. Stoica, *System identification*. Upper Saddle River, NJ: Prentice Hall, 1989.

[51] L. Ljung, *System identification - theory for the user (2nd edition)*. Upper Saddle River, NJ: Prentice Hall, 1999.

[52] T. Kailath, A. H. Sayed, and B. Hassibi, *Linear estimation*. Englewood Cliffs, NJ: Prentice Hall, 1999.

[53] P. Van Overschee and B. De Moor, *Subspace identification for linear systems: theory - implementation - applications*. Dordrecht, The Netherlands: Kluwer, 1996.

[54] T. Katayama, *Subspace methods for system identification*. London, U.K.: Springer, 2006.

[55] M. Verhaegen and V. Verdult, *Filtering and system identification: a least squares approach*. Cambridge, U.K.: Cambridge Univ. Press, 2007.

[56] A. Lindquist and G. Picci, *Linear Stochastic Systems*. Berlin, Germany: Springer-Verlag, 2015.

[57] L. A. Zadeh and C. A. Desoer, *Linear system theory: the state space approach*. New York: McGraw-Hill, 1963.

[58] G. Scarciotti and A. Astolfi, "Model reduction by matching the steady-state response of explicit signal generators," *IEEE Trans. Autom. Control*, vol. 61, no. 7, pp. 1995–2000, 2016.

[59] K. A. Gallivan, A. Vandendorpe, and P. Van Dooren, "Sylvester equations and projection-based model reduction," *J. Comp. Applied Math.*, vol. 162, no. 1, pp. 213–229, 2004.

[60] ——, "Model reduction and the solution of Sylvester equations," in *Proc. MTNS*, Kyoto, Japan, 2006.

[61] A. Astolfi, "Model reduction by moment matching for linear and nonlinear systems," *IEEE Trans. Autom. Control*, vol. 55, no. 10, pp. 2321–2336, 2010.

[62] ——, "Model reduction by moment matching, steady-state response and projections," in *Proc. 49th Conf. Decision Control*. IEEE, 2010, pp. 5344–5349.

[63] G. Scarciotti and A. Astolfi, "Data-driven model reduction by moment matching for linear and nonlinear systems," *Automatica*, vol. 79, pp. 340–351, 2017.

[64] J. Carr, *Application of center manifold theory*. New York: Springer-Verlag, 1981.

[65] C. I. Byrnes and A. Isidori, "Steady state response, separation principle and the output regulation of nonlinear systems," in *Proc. 28th Conf. Decision Control*, Tampa, FL, USA, 1989, pp. 2247–2251.

[66] A. Isidori and C. I. Byrnes, "Output regulation of nonlinear systems," *IEEE Trans. Autom. Control*, vol. 35, no. 2, pp. 131–140, 1990.

[67] B. V. Gnedenko and A. N. Kolmogorov, *Limit distributions for sums of independent random variables (2nd edition)*. Cambridge, MA: Addison-Wesley, 1968.

[68] B. S. Everitt and D. J. Hand, *Finite mixture distributions*. London, U.K.: Chapman and Hall, 1981.

[69] S. Asmussen, *Applied probability and queues*. New York: Springer, 2008.

[70] R. G. Gallager, *Stochastic processes: theory for applications*. Cambridge, U.K.: Cambridge Univ. Press, 2013.
[71] A. Padoan and A. Astolfi, "Moments of random variables: a linear systems-theoretic view," in *Amer. Control Conf.*, Seattle, WA, USA, 2017, pp. 1035–1040.
[72] G. Scarciotti and A. Astolfi, "Model reduction of neutral linear and nonlinear time-invariant time-delay systems with discrete and distributed delays," *IEEE Trans. Autom. Control*, vol. 61, no. 6, pp. 1438–1451, 2016.
[73] G. Grimmett and D. Welsh, *Probability: an introduction*. Oxford, U.K.: Oxford Univ. Press, 2014.
[74] R. B. Bapat, *Graphs and matrices*. London, U.K.: Springer, 2010.
[75] A. Isidori and C. I. Byrnes, "Steady-state behaviors in nonlinear systems with an application to robust disturbance rejection," *Ann. Rev. Control*, vol. 32, no. 1, pp. 1–16, 2008.
[76] A. Isidori, *Nonlinear control systems (3rd edition)*. New York: Springer-Verlag, 1995.
[77] G. Scarciotti and A. Astolfi, "Characterization of the moments of a linear system driven by explicit signal generators," *Proc. Amer. Control Conf.*, Chicago, IL, USA, 2015, pp. 589–594.
[78] J. K. Hale, *Ordinary differential equations (2nd edition)*. Malabar, FL, USA: Kreiger, 1980.
[79] R. E. Kalman, P. L. Falb, and M. A. Arbib, *Topics in mathematical system theory*. New York: McGraw-Hill, 1969, vol. 33.
[80] T. Kailath, *Linear systems*. Englewood Cliffs, NJ: Prentice Hall, 1980.
[81] E. D. Sontag, *Mathematical control theory: deterministic finite dimensional systems (2nd edition)*. New York: Springer-Verlag, 1998.
[82] P. Embrechts and H. Schmidli, "Modelling of extremal events in insurance and finance," *ZOR*, vol. 39, no. 1, pp. 1–34, 1994.
[83] A. V. Oppenheim, A. S. Willsky, and S. H. Nawab, *Signals and systems*. Upper Saddle River, N.J.: Prentice Hall, 1997.
[84] J. B. Conway, *Functions of one complex variable*. New York: Springer-Verlag, 1973.
[85] A. Padoan and A. Astolfi, "Model reduction by moment matching at isolated singularities for linear systems: a complex analytic approach," in *IFAC World Congr.*, Toulouse, France, 2017, pp. 6525–6528.
[86] ——, "Model reduction by moment matching at isolated singularities for linear systems: a geometric approach," in *Proc. 56th Conf. Decision Control*, Melbourne, Australia, 2017, pp. 4807–4812.
[87] ——, "Isolated singularities and moments of linear and nonlinear systems," *IEEE Trans. Autom. Control*, 2017, under review.
[88] T. J. Rothenberg, "Identification in parametric models," *Econometrica*, no. 3, pp. 577–591, 1971.
[89] F. W. Steutel and K. Van Harn, *Infinite divisibility of probability distributions on the real line*. New York: Marcel Dekker, 2003.
[90] E. Lukacs, *Characteristic functions (2nd edition)*. London, U.K.: Griffin, 1970.
[91] C. T. Chen and C. Desoer, "Controllability and observability of composite systems," *IEEE Trans. Autom. Control*, vol. 12, no. 4, pp. 402–409, 1967.
[92] S. Frühwirth-Schnatter, *Finite mixture and Markov switching models*. New York: Springer, 2006.
[93] W. L. Smith, "Renewal theory and its ramifications," *J. Roy. Statist. Soc.*, pp. 243–302, 1958.
[94] W. J. Stewart, *Probability, Markov chains, queues, and simulation: the mathematical basis of performance modeling*. Princeton, NJ: Princeton Univ. Press, 2009.
[95] M. Harchol-Balter, *Performance modeling and design of computer systems: queueing theory in action*. Cambridge, U.K.: Cambridge Univ. Press, 2013.
[96] G. Kotsalis and M. Dahleh, "Model reduction of irreducible markov chains," in *Proc. 42nd Conf. Decision Control*, Maui, Hawaii, USA, 2003, pp. 5727–5728.
[97] T. Runolfsson and Y. Ma, "Model reduction of nonreversible markov chains," in *Proc. 46th Conf. Decision Control*, New Orleans, LA, USA, 2007, pp. 3739–3744.
[98] K. Deng and D. Huang, "Model reduction of markov chains via low-rank approximation," in *Proc. Amer. Control Conf.*, Montreal, Canada, 2012, pp. 2651–2656.
[99] K. Ogata, *Discrete-time control systems*. Englewood Cliffs, NJ: Prentice Hall, 1995.
[100] A. J. Van Der Schaft and J. M. Schumacher, *An introduction to hybrid dynamical systems*. London, U.K.: Springer-Verlag, 2000.
[101] R. Goebel, R. G. Sanfelice, and A. R. Teel, "Hybrid dynamical systems," *IEEE Control Syst. Mag.*, vol. 29, no. 2, pp. 28–93, 2009.
[102] R. F. Curtain and H. Zwart, *An introduction to infinite-dimensional linear systems theory*. New York: Springer-Verlag, 1995.
[103] M. Tucsnak and G. Weiss, *Observation and control for operator semigroups*. Basel, Switzerland: Birkhäuser, 2009.
[104] N. K. Bose, *Multidimensional systems theory and applications (2nd ed.)*. Amsterdam, The Netherlands: Kluwer, 2003.



Alberto Padoan was born in Ferrara, Italy, in 1989. He received the Laurea Magistrale degree (M.Sc. equivalent) *cum laude* in automation engineering from the University of Padova, Italy, in 2013. He then joined the Control and Power Group at the Electrical and Electronic Engineering Department, Imperial College London, U.K., from which he received the Ph.D. degree in 2018.

He is currently a Research Associate in the Control Group, Department of Engineering, University of Cambridge, U.K., and a Research Associate of Sidney Sussex College. His research interests are focused on dynamical systems and control theory, with special emphasis on modelling, system identification and model reduction.



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 of 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 of 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), Northeastern University (2013).

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 150 journal papers, of 30 book chapters and of over 240 papers in refereed conference proceedings. He is the author (with D. Karagiannis and R. Ortega) of the monograph Nonlinear and Adaptive Control with Applications (Springer-Verlag).

He is the recipient of the IEEE CSS A. Ruberti Young Researcher Prize (2007), the IEEE RAS Googol Best New Application Paper Award (2009), the IEEE CSS George S. Axelby Outstanding Paper Award (2012), the Automatica Best Paper Award (2017). He is a Distinguished Member of the IEEE CSS, IET Fellow, IEEE Fellow and IFAC Fellow. He is the recipient of the 2015 Sir Harold Hartley Medal, for outstanding contribution to the technology of measurement and control, from the Institute of Measurement and Control.

He served as Associate Editor for Automatica, Systems and Control Letters, the IEEE Trans. on Automatic Control, the International Journal of Control, the European Journal of Control and the Journal of the Franklin Institute; as Area Editor for the Int. J. of Adaptive Control and Signal Processing; as Senior Editor for the IEEE Trans. on Automatic Control; and as Editor-in-Chief for the European Journal of Control. He is currently Editor-in-Chief of the IEEE Trans. on Automatic Control. He served as Chair of the IEEE CSS Conference Editorial Board (2010-2017) and in the IPC of several international conferences. He is a Member of the IEEE Fellow Committee (2016).