(This is a sample cover image for this issue. The actual cover is not yet available at this time.)

Journal of Computational and Applied Mathematics 236 (2012) 4193–4213

Contents lists available at SciVerse ScienceDirect

Journal of Computational and Applied Mathematics

journal homepage: www.elsevier.com/locate/cam

The truncated Stieltjes moment problem solved by using kernel density functions

P.N. Gavriliadis *, G.A. Athanassoulis

Department of Naval Architecture and Marine Engineering, National Technical University of Athens, Heroon Polytechniou 9, GR-157 73 Zografos, Athens, Greece

ARTICLE INFO

Article history:
Received 4 July 2010
Received in revised form 6 September 2011

MSC:
65R32
65F22
45Q05

Keywords:
Truncated Stieltjes moment problem
Kernel density function
Probability density function
Ill-posed problem
Moment asymptotics
Tail behavior

ABSTRACT

In this work the problem of the approximate numerical determination of a semi-infinite supported, continuous probability density function (pdf) from a finite number of its moments is addressed. The target space is carefully defined and an approximation theorem is proved, establishing that the set of all convex superpositions of appropriate Kernel Density Functions (KDFs) is dense in this space. A solution algorithm is provided, based on the established approximate representation of the target pdf and the exploitation of some theoretical results concerning moment sequence asymptotics. The solution algorithm also permits us to recover the tail behavior of the target pdf and incorporate this information in our solution. A parsimonious formulation of the proposed solution procedure, based on a novel sequentially adaptive scheme is developed, enabling a very efficient moment data inversion. The whole methodology is fully illustrated by numerical examples.

1. Introduction

In many problems in Physics, Technology and Finance, modeled in a probabilistic way, it is much easier to calculate the moments of the underlying probability measures than the probability measures (distributions) themselves. Thus, it has been an important question to understand how partial information contained in a number of moments can be used to approximately reconstruct the corresponding (unknown) probability measure. This is, in essence, the *truncated moment problem* of mathematical analysis, originated by Stieltjes in his paper "Recherchers sur les fractions continues" [1], and simply stated as follows:

"Recover a function $f(x)$, given its moments $\mu_n = \int x^n f(x)\, dx, \quad n = 0, 1, \ldots, N$".

The moment problem, in the univariate case, is naturally subdivided in three different subproblems, in accordance with the support of the target pdf: the *Hausdorff moment problem* ($supp(f) = [a, b]$), the *Stieltjes moment problem* ($supp(f) = [0, \infty)$), and the *Hamburger moment problem* ($supp(f) = (-\infty, \infty)$). The three subproblems exhibit both

* Corresponding author.
E-mail addresses: pmath@central.ntua.gr, takis.pmath@gmail.com (P.N. Gavriliadis), mathan@central.ntua.gr (G.A. Athanassoulis).

0377-0427/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.cam.2012.05.015

common features and differences. In this paper we shall focus on the case of continuously distributed probability measures, supported by the half-axis $[0, \infty)$, i.e. the *Stieltjes moment problem*.

Theoretically, various necessary and sufficient conditions under which the Stieltjes moment problem has a solution, either unique or not, can be found in the mathematical literature; see, e.g., the special monographs of [2–4] and the recent works [5–11], containing very interesting results on the Stieltjes moment problem. These results have been applied to several problems arising in mathematics, physics and economics; see e.g., [12–19]. The applied problem of the recovery of a semi-infinite supported pdf from moment data plays a key role in diverse scientific fields, ranging from *molecular theory* [20,21] to *cosmology* [22], *stochastic mechanics* [23–25], *quantum mechanics* [26–28], *chemical engineering* [29–31], *finance and economics* [32–35].

Numerically, the moment problem is universally recognized as a difficult *inverse problem* which leads to the solution of highly ill-posed systems of equations [36–38]. Thus, a fundamental question is how to reformulate the numerical moment problem in a way permitting the reliable and efficient determination (approximation) of the underlying pdfs.

In accordance with the general principles concerning the regularization of ill-posed problems (see, e.g., [36,39,40,37]), the following two points have been found to be of fundamental importance in implementing a well-posed numerical solution scheme to the moment problem:

(i) the choice of an appropriate representation for the function to be recovered, and
(ii) the use of appropriate restrictions on the (approximate) moment data, ensuring that the given moment sequence $\mu_1, \mu_2, \ldots, \mu_N$ does belong to the appropriate moment space.

In this way, it is possible to exploit much (if not all) of the existing *a priori* knowledge about the unknown pdf, resulting in regularized numerical solution schemes. Some properties of the target pdf that should be built in the solution space *a priori* (before starting the numerical procedure) in order to improve the numerical solution are: positivity, boundary conditions, tail behavior, modality information [41] and the knowledge about the "localization" of the target pdf [42] (i.e. the identification of the main-mass interval).

For a globally satisfactory, and at the same time "detailed", solution to the Stieltjes moment problem it is important to recover both the main-mass distribution and the tail behavior. The latter feature, indicating the main difference between the Stieltjes and the Hausdorff moment problems, has not been treated up to now, to the best of our knowledge. Existing methods treating the pdf-from-moments reconstruction either disregard (truncate) the tail, eventually solving a Hausdorff moment problem, or they are restricted to prespecified tail behavior (see e.g. [43,44]), lacking the flexibility to fit the "correct" tail behavior of the target pdf, as dictated by the moment sequence. It is known, however, that "*the asymptotic behavior of the moments is important for determining the tails of the probability density function*" [45]. Abate et al. (1999) proved a quantitative result relating the tail behavior with the asymptotic behavior of the moment sequence. In the present paper, this connection is exploited, in an efficient way, in order to dig out the "hidden" information leading to a rational tail selection on the basis of moment data. Although, formally speaking, these results concern high-order moment sequence asymptotics, it has been found in this work that 10 moments are enough to provide all the necessary information in order to rationally approximate the tail form.

The method proposed herewith for solving the Stieltjes moment problem, called the *K*ernel *D*ensity *E*lement *M*ethod (*KDEM*), has the following main features:

- It controls the admissibility of the moment data
- It is able to exploit asymptotic (moment) results in order to pick out the correct tail behavior and incorporate it in the approximate solution
- It is based on a positivity-preserving kernel-type representation
- It takes into account the support of the pdf to be approximated
- Its numerical solution is obtained by solving a constraint optimization problem
- It allows improvement by means of a sequentially adaptive scheme.

The structure of the main part of this paper is as follows. In Section 2, some aspects of the theoretical background of the Stieltjes moment problem are presented. In Section 3, the function space where the target pdfs "live in" (i.e. the solution space) is defined. It contains continuous pdfs having a parametric tail behavior with "free" tail parameters (to be determined during the solution procedure). In this connection the moment-sequence asymptotics is studied and related with the parameters of the target pdfs. In Section 4 an approximation theorem is proved, establishing that the set of all convex superpositions of appropriate kernel density functions is dense in the target (solution) space. Section 5 presents the implementation of the new moment-based method (KDEM) for approximating pdfs. The steps of the proposed reconstruction procedure are fully analyzed in an example. In Section 6, a parsimonious formulation for the proposed representation based on a novel, iterative, adaptive scheme, is presented and its performance is studied by an example. Finally, concluding remarks and indications for further developments are given in Section 7.

List of abbreviations and basic notation.

pdf: probability density function	KDF: Kernel Density Function
cdf: cumulative distribution function	KDR: Kernel Density Representation
ccdf: complementary cumulative distribution function	KDEM: Kernel Density Element Method
gG: generalized Gamma	SAA: Sequentially Adaptive Algorithm
X, Y: random variables	K: Kernel Density Function (KDF)
f, f_X, f_Y: probability density functions	K^{gG}: generalized Gamma KDF
$f_X^{(I)}, f_Y^{(I)}$: reconstructed (approximate) pdf	x_i, y_i: location parameter of the ith KDF
f_X^{gG}: generalized Gamma pdf	h: bandwidth parameter of KDF
$\mu_n, \mu_n^{(X)}, \mu_n^{(Y)}$: nth-order moments	d: tail-form parameter of pdf and KDF
$\mu_{n,exact}^{(X)}$: nth-order exact moments	p_i: probability weight of the ith KDF
$\hat{\mu}_n, \hat{\mu}_{n,k}$: nth-order approximate moments	I: index of complexity (number of KDFs)
N: total number of moments	$B_{n,i}$: nth-order moment of the ith KDF
$T = \bigcup_{d>0} \tilde{C}_0([0, +\infty); d)$: solution or target space	s: main-mass delimiter

2. The Stieltjes moment space $M^{(N)}$

This section is a brief overview of some known results, which are important for the correct formulation of the Stieltjes moment problem, and will be used in the numerical implementation (to define the admissible data sets; see Section 5.1 and Example 2). We shall restrict ourselves only to the *truncated Stieltjes problem*, i.e. the recovery of a density function $f(x)$ if we know only a finite set of its moments, say $\boldsymbol{\mu}_N = (\mu_1, \mu_2, \ldots, \mu_N)$. A given N-dimensional vector (point) $\boldsymbol{\mu}_N$ may or may not correspond to the N first moments of some pdf supported by $[0, +\infty)$. The set of all points $\boldsymbol{\mu}_N \in \mathbb{R}^N$, for which there exists at least one pdf having $\mu_1, \mu_2, \ldots, \mu_N$, as its first, second, ..., Nth-order moments, respectively, is called the (Nth-order) *moment space* $M^{(N)}$. The moment space $M^{(N)}$ is a *convex* subset of the non-negative orthant of $\mathbb{R}^N$, not necessarily closed or bounded [46].

If the point $\boldsymbol{\mu}_N$ belongs to the interior of the moment space $M^{(N)}$ (denoted by $\operatorname{Int} M^{(N)}$), the truncated moment problem is *indeterminate*, i.e., it has infinitely many solutions. If the point $\boldsymbol{\mu}_N$ belongs to the boundary of the moment space $M^{(N)}$ (denoted by $\partial M^{(N)}$), the truncated moment problem is *determinate*, i.e. it has a unique solution, which is represented by a discrete probability distribution [2,3].

Exact characterization of the Stieltjes moment space $M^{(N)}$ can be given in terms of the Hankel determinants:

$$H_{2s}^{(0)} := \begin{vmatrix} \mu_0 & \cdots & \mu_s \\ \vdots & & \vdots \\ \mu_s & \cdots & \mu_{2s} \end{vmatrix}, \qquad H_{2s+1}^{(1)} := \begin{vmatrix} \mu_1 & \cdots & \mu_{s+1} \\ \vdots & & \vdots \\ \mu_{s+1} & \cdots & \mu_{2s+1} \end{vmatrix}.$$

In fact, a necessary and sufficient condition for a given vector $\boldsymbol{\mu}_N$, lying in the non-negative orthant of $\mathbb{R}^N$, to belong to the Stieltjes moment space $M^{(N)}$ is that the following inequalities hold true:

$$H_{2s}^{(0)} > 0, \quad \text{for all } s,\ 2 \le 2s < N, \quad \text{and} \tag{1a}$$

$$H_{2s+1}^{(1)} > 0, \quad \text{for all } s,\ 3 \le 2s+1 < N. \tag{1b}$$

The proof of this result can be found in [47] or [3].

3. The solution space T

In this section, an appropriate *admissible solution space* for the target pdfs supported by the semi-infinite interval $[0, +\infty)$, is defined in terms of its boundary behavior. In this way, the theoretical and numerical connection between the moment sequence asymptotics and the tail behavior of the target pdf is established.

3.1. Definition of the solution space

The first restriction we impose to the solution space is the assumption that *all moments exist*, a natural assumption in the context of the moment problem.

Since in many engineering and environmental applications, $\mathbb{R}^+$-supported pdfs are assumed to have a zero value at $x = 0$, we adopt this assumption in our solution space, i.e.

(i) $f(x) = 0, \quad \text{at } x = 0.$

Furthermore, we assume that all considered target pdfs have a tail of the form (generalized Gamma tail, *gG*-tail)

$$\text{(ii)}\quad f(x) \sim ax^{-b}\exp\left\{-cx^{d}\right\}, \quad \text{as } x\to\infty,\ (a,c,d>0,\ b\in\mathbb{R}), \tag{2a}$$

where $f(x)\sim g(x)$ means that $f(x)/g(x)\to 1$ as $x\to\infty$, or, more precisely,

$$(\forall\varepsilon>0)\ (\exists M(\varepsilon)>0)\ (\forall x\geq M(\varepsilon)) : \left|f(x)-ax^{-b}\exp\left\{-cx^{d}\right\}\right|<\varepsilon. \tag{2b}$$

The tail form (2) contains the exponential ($f(x)\sim a\exp\{-cx\}$), the semi-exponential ($f(x)\sim ax^{-b}\exp\{-cx\}$) and the maximum entropy tails as special cases. The parameters a, b, c and d appearing in (2) will be called *tail form parameters*. Of special importance is the parameter d, which controls the rate of exponential decay of f.

The set of all continuous pdfs satisfying the boundary conditions (i) and (ii) for a specific value of $d>0$, will be denoted by $\tilde{C}_0([0,+\infty);d)$. In this work, a target pdf will be looked for in the set

$$T=\bigcup_{d>0}\tilde{C}_0([0,+\infty);d), \tag{3}$$

which will be referred to as the *solution* or *target space*. It can be easily proved that T is a convex metric space under the distance $d(f,g)=\sup_{x\in[0,+\infty)}|f(x)-g(x)|$.

Modifications of the solution space (e.g., by using different boundary conditions either at $x=0$ or at $x\to\infty$) are possible. In such a case, the numerical procedure presented below should be modified appropriately.

It can be easily proved that any element of the target space T is a *bounded* and *uniformly continuous* function over the infinite interval $[0,+\infty)$. The target space T is wide enough, so that it can provide solutions to a great number of problems and, at the same time, permits us to construct an explicit approximate representation of its elements (Section 4), incorporating the tail behavior of the target pdf (the parameter d) by exploiting asymptotic properties for the moments, presented in the next subsection.

3.2. The tail of the target pdf and the moment sequence asymptotics

Under the assumption (2) it is possible to obtain an asymptotic expansion of the moment sequence $\{\mu_n\}$, dependent only on the tail form parameters c and d, as follows:

$$\mu_n\sim A\frac{\Gamma[((n+B)/d)+1]}{c^{(n+B)/d}}, \quad \text{as } n\to\infty, \tag{4}$$

where A, B are constants dependent on a, b but not on n. This asymptotic result has been established by Abate et al. (1999).

The dependence of μ_n on n is clarified further by using the following enhanced version of the well-known Stirling formula for the Gamma function (see Eq. (44) in [48]),

$$\Gamma[az+b]=\sqrt{2\pi}\,e^{-az}(az)^{az+b-1/2}\left(1+\frac{g_1(b)}{az}+\cdots+\frac{g_{m-1}(b)}{(az)^{m-1}}+O\left((az)^{-m}\right)\right), \quad \text{as } z\to\infty, \tag{5}$$

where $g_m(b)$ are functions dependent only on b. The first two of them are given by formulae $g_1(b)=\left(6b^2-6b+1\right)/12$, $g_2(b)=\left(36b^4-120b^3+120b^2+36b+1\right)/288$. Combining (4) and (5) we obtain

$$\mu_n\sim A\sqrt{2\pi}\,c^{(n+B)/d}e^{-n/d}\left(\frac{n}{d}\right)^{((n+B)/d)+\frac{1}{2}}\left(1+\frac{\lambda_1}{n}+\cdots+\frac{\lambda_{m-1}}{n^{m-1}}+O\left(n^{-m}\right)\right), \quad \text{as } n\to\infty, \tag{6}$$

where λ_m are dependent on d but not on n.

The asymptotic expansion (6) will now be exploited in order to approximately determine the tail parameters c and d from moment asymptotics. To proceed with, let us consider the ratios

$$r_n:=\frac{\mu_n}{\mu_{n-1}}, \qquad d_n:=\frac{r_n}{n(r_{n+1}-r_n)} \quad \text{and} \quad c_n:=\frac{n}{d(r_n)^d}. \tag{7}$$

It can be shown (see Theorem 6.1 in [49]) that

$$\lim_{n\to\infty}d_n=d \quad \text{and} \quad \lim_{n\to\infty}c_n=c. \tag{8}$$

Table 1
The initial sequences c_n and d_n along with the $(n-1)$-th order approximations $\tilde{d}_n\,(n-1)$ and $\tilde{c}_n\,(n-1)$, for $n=5,\ldots,12$ (the case shown in Fig. 1).

n	c_n	$\tilde{c}_n\,(n-1)$	d_n	$\tilde{d}_n\,(n-1)$
5	1.4296840534	2.5865626661	8.7825844741	3.9789586220
7	1.6861081285	2.9197650731	7.4074808076	3.9946490178
10	1.9481675714	2.9963911805	6.3788462027	3.9996889523
12	2.0735688726	2.9996656160	5.9797683734	3.9999788544

Furthermore, by using asymptotic expansion (6), it is straightforward to show that both sequences c_n and d_n converge to the exact values c and d with a slow rate $O\left(n^{-1}\right)$, i.e.

$$d_n = d + \sigma_1 n^{-1} + \sigma_2 n^{-2} + \cdots \quad \text{and} \quad c_n = c + \tau_1 n^{-1} + \tau_2 n^{-2} + \cdots \text{ as } n \to \infty, \tag{9}$$

where $(\sigma_1, \sigma_2, \ldots)$ and $(\tau_1, \tau_2, \ldots)$ are n-independent quantities. This slow convergence rate seems to preclude the possibility of exploiting the above asymptotic result for the determination of the parameters c and d, in the context of the truncated moment problem (where only a small number of moments are usually known). The situation is changing dramatically, however, if we exploit a Richardson extrapolation scheme [49,50], which leads to nice convergence acceleration results. The Richardson extrapolation scheme is formulated as follows:

Consider first the k equations (derived from (9))

$$\begin{aligned} d_{n+i} &= d + \sigma_1 (n+i)^{-1} + \sigma_2 (n+i)^{-2} \\ &\quad + \cdots + \sigma_{k-1} (n+i)^{-(k-1)} + O\left((n+i)^{-k}\right), \quad \text{for } i = 0, 1, \ldots, k-1, \end{aligned} \tag{10a}$$

and then truncate system (10a), retaining the k first terms

$$d_{n+i} = \tilde{d}_n\,(k) + \sigma_1 (n+i)^{-1} + \sigma_2 (n+i)^{-2} + \cdots + \sigma_{k-1} (n+i)^{-(k-1)}, \quad \text{for } i = 0, 1, \ldots, k-1. \tag{10b}$$

Solving this system, we obtain a closed-form expression for $\tilde{d}_i\,(k)$ in terms of d_{n+i}, $i = 0, 1, \ldots, k-1$ (called the k-th *order approximation* of d; see Eq. (8.1.16) in [50] or [49]):

$$\tilde{d}_n\,(k) = \frac{n^{k-1}}{(k-1)!} d_n - \frac{(n-1)^{k-1}}{(k-2)!} d_{n-1} + \cdots + (-1)^{k-1} \frac{(n-k+1)^{k-1}}{(k-1)!} d_{n-(k-1)}, \tag{11}$$

with $\tilde{d}_n\,(1) = d_n$. The new sequence $\tilde{d}_i\,(k)$, for $i = 2, \ldots, n, \ldots, k = 1, \ldots, i-1$, converges to the same limit value d, with the highly improved rate $O\left(n^{-k}\right)$ [49]. The same acceleration procedure is also applied to sequence $c_1, c_2, \ldots, c_n, \ldots$. In some cases, it is more efficient to apply the extrapolation scheme to sequences $1/d_n$, $1/c_n$ instead of d_n, c_n.

To illustrate the practical significance of the Richardson extrapolation formulae (11), an elementary example is presented. A specific pdf is considered, for which the analytic form and the tail parameters and the moment sequence are explicitly known, so that it is easy to calculate the sequences c_n, d_n and $\tilde{c}_n$, $\tilde{d}_n$ and illustrate their convergence towards the exact values of c and d, respectively.

Example 1. Consider a generalized gamma distribution

$$f_X^{gG}(x; b, c, d) = \frac{dc^b}{\Gamma[b]} x^{bd-1} \exp\left\{-cx^d\right\}, \quad x \geq 0, (b, c, d > 0), \tag{12a}$$

having moments

$$\mu_{n,exact}^{(X)} = c^{-(n/d)} \frac{\Gamma[(n/d)+b]}{\Gamma[b]}, \quad n = 1, 2, \ldots. \tag{12b}$$

The choice of the parameters in our example (Fig. 1) is as follows: $b = 2$, $c = 3$, $d = 4$. First, we calculate the $N = 40$ first moments (Eq. (12b)) and then the sequences c_n and d_n (Eq. (7)) for $n = 5, \ldots, 39$. The two sequences c_n and d_n, as the number of moments increases, are presented in Fig. 1 (using stars). Their disappointing slow convergence is clearly seen in the figure (see also Table 1). Then, we calculate the $k = (n-1)$-th order approximations $\tilde{d}_n\,(n-1)$ and $\tilde{c}_n\,(n-1)$, for $n = 5, \ldots, 12$ (Eq. (11)), and plot the results in the same Fig. 1 (using small circles). It is clearly seen that the accelerated sequences $\tilde{d}_n\,(n-1)$ and $\tilde{c}_n\,(n-1)$ converge much faster, providing a very good accuracy by using only 5–7 moments ($\tilde{c}_7\,(6) = 2.9197650731$, $\tilde{d}_7\,(6) = 3.9946490178$; see also Table 1).

4. The Kernel Density Representation (KDR) of the target pdfs

In this section, we establish a convenient approximate representation of the elements of the target space T, in the form of convex superpositions of kernel density functions. The main result is Theorem 2, ensuring that the set of all convex superpositions of KDFs is dense in T.

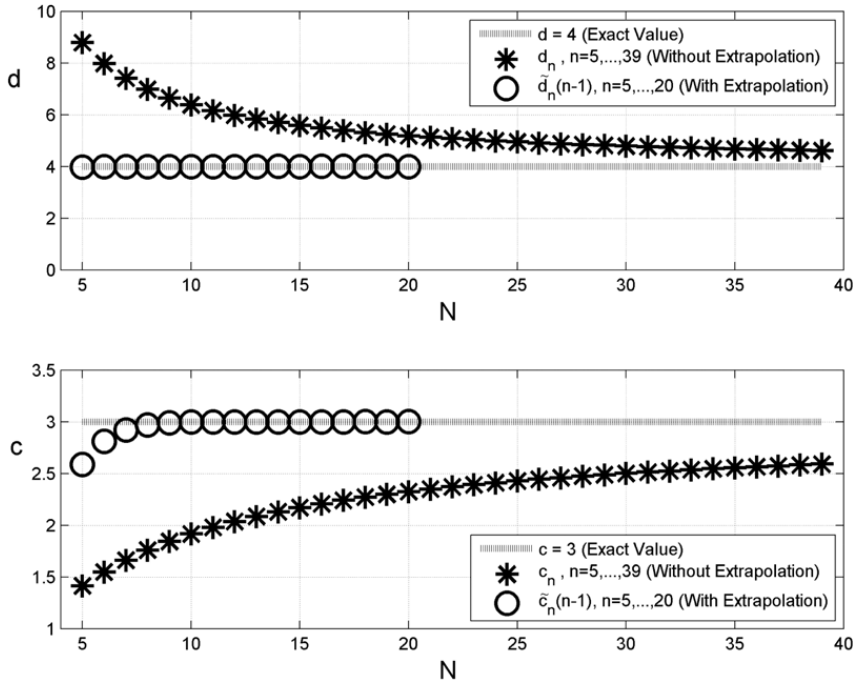

Fig. 1. Extrapolation of the tail form parameters d, c with respect to the number of moment data.

A natural constraint which will be implemented as an *a priori* condition of any approximate representation of an element in the solution space, is the *positivity* (*non-negativity*) constraint:

Definition 1. A representation $f\left(x; \mathbf{p}_{(I)}\right) = f\left(x; p_1, p_2, \ldots, p_I\right)$ of a pdf $f\left(x\right)$ will be called *positivity-preserving* iff it satisfies the following property

$$\mathbf{p}_{(I)} \geq \mathbf{0} \Rightarrow f\left(x; \mathbf{p}_{(I)}\right) \geq 0, \quad \text{for any } I = 1, 2, 3, \ldots. \tag{13}$$

The positivity-preserving property ensures that the recovered function will be a legitimate (bona fide) pdf, that is it will not become negative at any part of the real axis. This property is not inherent in general approximation or representation theorems. For example, any truncated Fourier series (or Fourier integral) representation of pdfs may (and usually does) become negative in some intervals of the real line [51,52]. There are only a few positivity-preserving representations of functions. Except from the kernel-based representation (see below, Eq. (14a)), we mention representations defined in terms of a linear combination of the product of positive polynomials with exponential functions [53], and various exponential-type representations (e.g., exponentials of orthogonal polynomials [54] and the famous Maximum Entropy representation [55,25,56,35]). A detailed review on the positivity-preserving representations can be found in [38].

A convenient candidate to serve as a positivity-preserving, approximate representation of pdfs, is the family of all *convex superpositions of Kernel Density Functions*:

$$f_X^{(I)}\left(x\right) = f\left(x; \mathbf{p}_{(I)}\right) := \sum_{i=1}^{I} p_i \, K\left(x; x_i, h, d\right), \quad x \in [0, +\infty), \tag{14a}$$

where

$K\left(x; x_i, h, d\right)$ ith *component KDF*; the specific choice made in the present work for the KDF is the generalized gamma function, which is explicitly given and discussed in Appendix A,
x_i *location parameter*; in this work x_i is chosen to be the mode of each KDF,
h *bandwidth parameter*; in this work h is chosen to be the standard deviation of each KDF,
d *tail-form parameter* associated with the limiting behavior of $f\left(x\right)$ as $x \to \infty$ (see Eq. (2)),
p_i *probability weight* of the ith KDF,
I *index of complexity*, i.e. the order of the mixture model.

The index I is supposed to be finite, while the coefficients p_i, $i = 1, 2, \ldots, I$, are always restricted to comply with the conditions:

$$p_i \geq 0, \quad i = 1, 2, \ldots, I \quad \text{and} \quad \sum_{i=1}^{I} p_i = 1. \tag{14b}$$

For fixed $I \in \mathbb{N}$, the *admissible domain* $Q^{(I)}$ of the p_i coefficients forms a simplex (convex compactum), defined by $Q^{(I)} := \left\{p_i, i = 1, \ldots, I : \sum_{i=1}^{I} p_i = 1, p_i \geq 0\right\} \subset \left(\mathbb{R}_0^+\right)^I$.

Remark 1. Note that the usual assumption $K(x; x_i, h, d) = K(x - x_i; h, d)$ is not imposed here, because it is not compatible with the condition $f(x) = 0$, for $x < 0$, which is an essential condition for the Stieltjes problem.

Remark 2. The location parameters $x_i, i = 1, \ldots, I$, are I points belonging to a finite subdomain $\hat{A} = [0, s]$ of $A = [0, +\infty)$. In this work they are chosen to be equi-distributed on $\hat{A} = [0, s]$. The choice of s will be discussed in Section 5.1. It should be emphasized that the specific value of s *does not* affect the tail form of the target pdf. The latter is determined by means of the tail form parameter d, which is incorporated in the KDFs and is calculated independently by using the moment asymptotics, developed in Section 3.2.

KDF form a special case of kernel functions (cf. *summability kernels* in harmonic analysis), commonly used for constructing representation theorems in Analysis [57]. In our case, KDF is assumed to be a specific appropriate element of the target space $T = \tilde{C}_0([0, +\infty); d)$. The essential properties of our kernel density functions are the following:

(i) $K(x; w, h, d)$ is a unimodal, uniformly continuous pdf, with location parameter (mode) w and bandwidth (standard deviation) h, defined in $D_K = A \times A \times B \times B$, where $A = [0, +\infty)$ and $B = (0, +\infty)$;
(ii) $K(x; w, h, d) \geq 0, \forall (x; w, h, d) \in D_K$;
(iii) $\int_A K(x; w, h, d)\, dx_i = 1, \forall (x; w, h, d) \in D_K$;
(iv) $(\forall \varepsilon > 0)\,(\forall \delta > 0)\left(\exists \tilde{h}(\varepsilon, \delta) > 0\right)\left(\forall h \in \left(0, \tilde{h}(\varepsilon, \delta)\right)\right) : \int_{|x-w|>\delta} K(x; w, h, d)\, dw < \varepsilon$, uniformly in $x \in A$, for any given $d \in B$.

For example, the Gamma, the generalized Gamma and the Chi-square density functions satisfy all the above properties, being appropriate candidates for KDFs.

The main result of this section, which will be established by means of the following two theorems, is that the set of all convex superpositions of KDFs satisfying properties (i)–(iv) form a dense set in the solution space T.

Theorem 1. *Let $f \in T = \tilde{C}_0([0, +\infty); d)$ be a pdf from the target space, and $K(\cdot; \cdot, \cdot, d)$ be any KDF satisfying properties* (i)–(iv)*, stated above. Then, for any $\varepsilon > 0$, there exists an $\tilde{h}(\varepsilon) > 0$, such that*

$$\left| \int_A K(x; w, h, d) f(w)\, dw - f(x) \right| < \varepsilon, \quad \text{uniformly in } x \in A, \tag{15}$$

provided that $0 < h < \tilde{h}(\varepsilon)$.

Proof. The proof is based on the following decomposition of the difference between $\int_A K(x; w, h, d) f(w)\, dw$ and $f(x)$:

$$\begin{aligned}
&\left| \int_A K(x; w, h, d) f(w)\, dw - f(x) \right| \\
&\leq \int_0^{+\infty} K(x; w, h, d) \left| f(w) - f(x) \right| dw \\
&= \int_0^{x-\delta} K(x; w, h, d) \left| f(w) - f(x) \right| dw + \int_{x-\delta}^{x+\delta} K(x; w, h, d) \left| f(w) - f(x) \right| dw \\
&\quad + \int_{x+\delta}^{+\infty} K(x; w, h, d) \left| f(w) - f(x) \right| dw = I_{0,x-\delta} + I_{x-\delta,x+\delta} + I_{x+\delta,+\infty}.
\end{aligned} \tag{16}$$

In order to estimate the various terms appearing in the r.h.s of the above inequality, some preliminary inequalities are first introduced, coming from the aforementioned properties of the target space (see Section 3.1) and the KDFs (properties (i)–(iv), above).

(a) Since f is uniformly continuous in $[0, +\infty)$, given $\varepsilon > 0$, there exists a $\delta(\varepsilon) > 0$ such that

$$|x - w| < \delta(\varepsilon) \Rightarrow |f(x) - f(w)| < \frac{\varepsilon}{3}.$$

(b) Since f is bounded in $[0, +\infty)$, for every $\delta > 0$ there exists $M_0 > 0$ such that

$$\begin{aligned}
&|f(x) - f(w)| < M_0, \quad \forall w \in [0, x - \delta], \quad \text{and} \\
&|f(x) - f(w)| < M_0, \quad \forall w \in [x + \delta, +\infty).
\end{aligned}$$

(c) Keeping the same $\varepsilon > 0$, choosing $\delta = \delta(\varepsilon) > 0$ and M_0 as defined in (a) and (b) above, and invoking property (iv) of KDFs, there exists an $\tilde{h}(\varepsilon, \delta(\varepsilon)) > 0$ such that

$$\int_{|x-w|>\delta(\varepsilon)} K(x; w, h, d)\, dw < \frac{\varepsilon}{3M_0}, \quad \text{for } 0 < h < \tilde{h}(\varepsilon, \delta(\varepsilon)),$$

which, in conjunction with property (ii), implies that

$$\int_0^{x-\delta(\varepsilon)} K(x; w, h, d)\, dw < \frac{\varepsilon}{3M_0}, \quad \text{and} \quad \int_{x+\delta(\varepsilon)}^{+\infty} K(x; w, h, d)\, dw < \frac{\varepsilon}{3M_0}.$$

On the basis of the inequalities appearing in (b) and (c), above, the integrals $I_{0,x-\delta(\varepsilon)}$ and $I_{x+\delta(\varepsilon),+\infty}$ can be estimated as follows:

$$I_{0,x-\delta(\varepsilon)} = \int_0^{x-\delta(\varepsilon)} K(x; w, h, d)\ |f(w) - f(x)|\, dw \le \frac{\varepsilon}{3M_0} \cdot M_0 = \frac{\varepsilon}{3}, \tag{17a}$$

$$I_{x+\delta(\varepsilon),+\infty} = \int_{x+\delta(\varepsilon)}^{+\infty} K(x; w, h, d)\ |f(w) - f(x)|\, dw \le \frac{\varepsilon}{3M_0} \cdot M_0 = \frac{\varepsilon}{3}. \tag{17b}$$

Furthermore, taking into account the inequality appearing in (a), the integral $I_{x-\delta(\varepsilon),x+\delta(\varepsilon)}$ can be estimated as follows:

$$I_{x-\delta(\varepsilon),x+\delta(\varepsilon)} = \int_{x-\delta(\varepsilon)}^{x+\delta(\varepsilon)} K(x; w, h, d)\ |f(w) - f(x)|\, dw \le \frac{\varepsilon}{3} \cdot \int_0^{+\infty} K(x; w, h, d) dw \le \frac{\varepsilon}{3} \cdot 1 = \frac{\varepsilon}{3}. \tag{17c}$$

Combining now inequalities (17a)–(17c) with (16), we obtain (15). This completes the proof of Theorem 1. □

Theorem 2. *Let $f \in T = \bigcup_{d>0} \tilde{C}_0([0, +\infty); d)$ and $K(\cdot; \cdot, \cdot, \cdot)$ be any KDF satisfying properties* (i)–(iv)*, stated above. Then, given any $\varepsilon > 0$, there exists a finite set $\mathbf{p} = (p_1, p_2, \ldots, p_I) \in Q^{(I)}$ of non-negative constants, a finite set of centers $\{x_i\}_{i=1}^I$ from $A = [0, +\infty)$, and an $\tilde{h}(\varepsilon) > 0$, such that*

$$\left| f(x) - \sum_{i=1}^{I} p_i K(x; x_i, h, d) \right| < \varepsilon, \quad \textit{uniformly in } x \in A, \tag{17}$$

provided that $0 < h < \tilde{h}(\varepsilon)$.

Proof. Since $f \in T = \bigcup_{d>0} \tilde{C}_0([0, +\infty); d)$, there exists a $d > 0$, such that $f \in \tilde{C}_0([0, +\infty); d)$. Choose a KDF $K(\cdot; \cdot, \cdot, d)$ with the same tail parameter d. Thus, any convex superposition $\sum_{i=1}^{I} p_i K(x; x_i, h_i, d)$ decays to zero as $x^{-b} \exp\{-cx^d\}$, when $x \to +\infty$.

On the basis of (15), given any $\varepsilon > 0$, there exists an $\tilde{h}(\varepsilon) > 0$ such that

$$\left| f(x) - \int_A K(x; w, h, d) f(w)\, dw \right| < \frac{\varepsilon}{2}, \tag{18a}$$

provided that $h < \tilde{h}(\varepsilon)$. In addition, the following two properties hold true:

(a) Given any $\varepsilon > 0$, there exists an $M(\varepsilon) > 0$, such that

$$\left| \int_{M(\varepsilon)}^{+\infty} K(x; w, h, d) f(w)\, dw \right| \le \frac{\varepsilon}{4}, \quad \text{for all } x \in [M(\varepsilon), +\infty).$$

(b) For the same $M(\varepsilon) > 0$, as defined above, the integral of $K \cdot f$ over the finite domain $[0, M(\varepsilon)]$ can be uniformly approximated by its Riemann sum. That is, for the same $\varepsilon > 0$, as in (a), it is possible to find $I \in \mathbb{N}$ and $x_i \in [0, M(\varepsilon)]$, $i = 1, \ldots, I$, such that

$$\left| \int_0^{M(\varepsilon)} K(x; w, h, d) f(w)\, dw - \sum_{i=1}^{I} p_i K(x; x_i, h, d) \right| < \frac{\varepsilon}{4}, \quad \text{uniformly in } x \in A,$$

where $p_i = f(x_i)\ \Delta x_i$, and $\Delta x_i = x_{i+1} - x_i$, $i = 1, \ldots, I$. Combining the above inequalities, we obtain:

$$\begin{aligned} &\left| \int_A K(x; w, h, d) f(w)\, dw - \sum_{i=1}^{I} p_i K(x; x_i, h, d) \right| \\ &= \left| \int_0^{M(\varepsilon)} K(x; w, h, d) f(w)\, dw + \int_{M(\varepsilon)}^{+\infty} K(x; w, h, d) f(w)\, dw - \sum_{i=1}^{I} p_i K(x; x_i, h, d) \right| \end{aligned}$$

$$\leq \left| \int_0^{M(\varepsilon)} K(x;\, w,\, h,\, d) f\,(w)\, dw - \sum_{\substack{i=1 \\ x_i \in [0, M(\varepsilon)]}}^{I} p_i K(x;\, x_i,\, h,\, d) \right|$$

$$+ \left| \int_{M(\varepsilon)}^{+\infty} K(x;\, w,\, h,\, d) f\,(w)\, dw \right| \leq \frac{\varepsilon}{4} + \frac{\varepsilon}{4} = \frac{\varepsilon}{2}. \tag{18b}$$

Combining now (18a) and (18b), we easily obtain inequality (17). This completes the proof of Theorem 2. □

5. Numerical solution of the moment problem. The KDEM

In this section we describe a numerical method to reconstruct pdfs from a number of their moments by using the KDR. The inherent ill-posedness of the moment problem is fully controlled by means of the *a priori* specification of the moment-data space (Section 2) and the KDR (Section 4). These tools are combined in this section with additional moment-based information, such as e.g., the "localization" of the target pdf, and a constrained optimization formulation, defining the proposed KDEM.

5.1. Data preprocessing

– *Numerical "admissibility" of the moment data*

The numerical "admissibility" of the moment data, i.e. whether they belong to the moment space or not, is checked and controlled by means of conditions (1a) and (1b) (see [58] for more details).

– *Rescaling to a computationally efficient interval*

For numerical purposes, i.e. avoiding to perform computations with numbers exhibiting a very large difference in their order of magnitude, the domain $A \equiv A^{(X)} = [0, +\infty)$, i.e. the sample space of the underlying random variable X, is subdivided into two subintervals, the *main-mass interval* $A^{(X)}_{main}$ and the *tail interval* $A^{(X)}_{tail}$:

$$A^{(X)} = [0, +\infty) = [0, s] \cup (s, +\infty) = A^{(X)}_{main} \cup A^{(X)}_{tail}, \tag{19}$$

where s, the *main-mass delimiter*, is defined by

$$s : F\,(s) > 1 - \varepsilon, \tag{20}$$

where ε is a small positive quantity. Two different approaches for the determination of the main-mass delimiter s, based on the knowledge of moments, have been recently proposed [42,59]. In the present work method [42] will be used. See Example 2.

After calculating the main-mass delimiter s, a linear transformation $Y = (l/s)\, X$ is applied, transforming the main-mass interval $A^{(X)}_{main} = [0, s]$ into the normalized interval $A^{(Y)}_{main} = [0, l]$, where $l \leq 1$ (usually $l \approx 0.8$). See Fig. 2. Since

$$f_Y\,(y) = (s/l)\, f_X\,(sy/l)\,, \tag{21a}$$

the moments are transformed in accordance with the equation

$$\mu^{(Y)}_n = (l/s)^n\, \mu^{(X)}_n, \quad n = 0, 1, \ldots, N. \tag{21b}$$

Remark 3. The transformation $Y = (l/s)\, X$ is defined on the basis of main-mass delimiter s, but it applies on the whole domain $A = [0, +\infty)$. That is, *it is a rescaling and not a tail-truncation* of the target pdf.

5.2. Constrained-optimization solution of the truncated moment problem

Multiplying both sides of Eq. (14a) by x^n, and integrating over $[0, +\infty)$, we obtain the following equations, connecting the parameters p_i, $i = 1, 2, \ldots, I$, and the moment data of the target pdf:

$$\mu_n = \sum_{i=1}^{I} B_{n,i}\, p_i, \quad n = 0, 1, \ldots, N, \quad \text{or, in matrix form,} \quad \boldsymbol{\mu} = \mathbf{B}\mathbf{p}, \tag{22a}$$

where

$$B_{n,i} = B_n\,(x_i) = \int_0^{\infty} x^n K\,(x;\, x_i,\, h,\, d)\;\, dx, \quad n = 0, 1, \ldots, N,\; i = 1, 2, \ldots, I, \tag{22b}$$

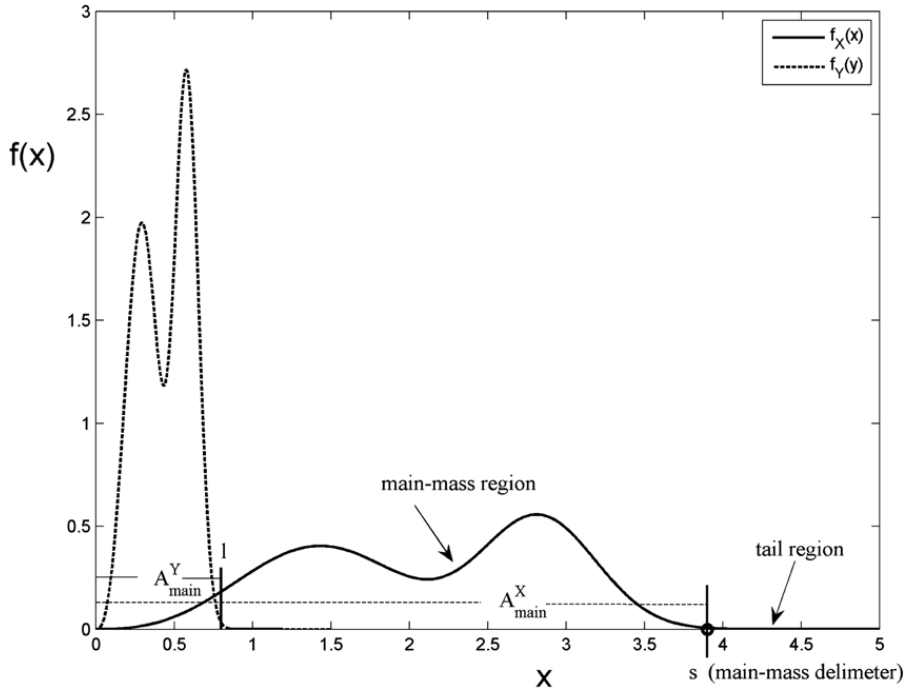

Fig. 2. Rescaling. Transformation of the main-mass interval $A^{(X)}_{main} = [0, s]$ into the normalized interval $A^{(Y)}_{main} = [0, l]$.

are the moments of the kernel density functions. Let us note that in all numerical examples which presented in this work, the generalized Gamma KDF is used (see Appendix A for details on its form and its moments).

As is well known, Problem (22) is a typical severely ill-conditioned inverse problem [36,37]. A key idea in reformulating problem (22) as a well-posed problem is to *a priori* restrict the moment data $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_N)$ to belong to the admissible moment space $M^{(N)}$ and the sought-for solution $\mathbf{p} = (p_1, p_2, \ldots, p_l)$ to belong to the appropriate compactum $Q^{(l)}$, defined by Eq. (14b). Thus, problem (22) is restated as a constrained optimization problem:

$$\text{Minimize } \Phi(\mathbf{p}) = \frac{1}{2}\|\boldsymbol{\mu} - \mathbf{B}\,\mathbf{p}\|^2, \quad \text{subject to } \mathbf{p} \in Q^{(l)}. \tag{23}$$

The constraint optimization problem (23) is numerically realized using the function *lsqlin* of the MATLAB optimization package [60,61].

5.3. Measures of numerical convergence

To track the numerical convergence of the proposed numerical scheme, three distances are introduced. The standard, *pdf-based*, maximum and L^1 distances, denoted by

$$d_{\max}(f, g) := \max\{|f(x) - g(x)|, x \in [0, +\infty)\}, \tag{24a}$$

$$d_1(f, g) := \int_0^{+\infty} |f(x) - g(x)|\, dx, \tag{24b}$$

and a specific, *moment-based*, *semi-distance*, defined by

$$d_{mom}\left(\{\mu_n\}_{n=1}^N, \{\hat{\mu}_n\}_{n=1}^N\right) := \frac{1}{N}\left[\sum_{n=1}^N \left(\frac{\mu_n - \hat{\mu}_n}{(1/2)(\mu_n + \hat{\mu}_n)}\right)^2\right]^{1/2}, \tag{25}$$

where $\{\mu_n\}$ is the set of given moments and $\{\hat{\mu}_n\}$ is the set of the corresponding moments of the approximate solution. The latter distance measures how well the moments of the reconstructed pdf fit to the moment data, and it is called *M*ean *R*elative *E*rror (MRE).

Remark 4. Let us note here that the *approximate moments* $\hat{\mu}_n := \int_0^\infty x^n f_X^{(l)}(x)\, dx, n = 1, \ldots, N$, of the approximate solution $f_X^{(l)}(x)$ are given by $\hat{\mu}_n = \sum_{i=1}^l p_i \int_0^\infty x^n K(x; x_i, h)\, dx = \sum_{i=1}^l p_i B_{i,n}, n = 1, \ldots, N$.

Table 2
The values of the initial and transformed moments ($\mu_n^{(X)}$ and $\mu_n^{(Y)}$), for $n = 1, \ldots, 16$.

n	1	2	3	4	5	6
$\mu_n^{(X)}$	2.1371071785	5.2174052327	13.8374999774	38.6983478663	1.1219630534×10^{2}	3.3397312208×10^{2}
$\mu_n^{(Y)}$	0.4382712764	0.2194259465	0.1193459676	0.0684477610	0.0406969707	0.0248434410
n	7	8	9	10	11	12
$\mu_n^{(X)}$	1.0149316141×10^{3}	3.1380763299×10^{3}	9.8501109361×10^{3}	3.1342266465×10^{4}	1.0098976892×10^{5}	3.2926252662×10^{5}
$\mu_n^{(Y)}$	0.0154829502	0.0098174149	0.0063196273	0.0041237986	0.0027249663	0.0018219765
n	13	14	15	16		
$\mu_n^{(X)}$	1.0855715335×10^{6}	3.6174917497×10^{6}	1.2178808992×10^{7}	4.1408494276×10^{7}		
$\mu_n^{(Y)}$	0.0012319004	$8.4186300894\times10^{-4}$	$5.8123987770\times10^{-4}$	$4.0528145140\times10^{-4}$		

The detailed, pdf-based, distances (24a) and (24b) can be used *only* when the target pdf is analytically known, that is *only* for validation purposes; see Example 2. In general, the target pdf is unknown, and thus the use of distances (24a) and (24b) is meaningless. In these cases, the moment-based, distance (25) is always used to evaluate the "goodness" of the solution and, in particular, to establish the numerical convergence of any iterative solution schemes (see Section 6). Let it be noted that some discrepancy ($d_{mom} \neq 0$) is expected to exist even for the final solution of the moment problem, since the number of equations is not equal to the number of unknowns, and the problem is solved in the mean-square sense.

5.4. An example

To illustrate the above described numerical method of solution of the moment problem, we will apply it to reconstruct a pdf from a small number of (inexact) moments. For this purpose, a mixture of generalized gamma distributions is selected as the target pdf, and it is reconstructed using KDR from its 16 moments. In this example, we allow the noise to come into the inversion procedure by means of the numerical calculation of the moments.

Example 2. Consider the two-component mixture

$$f_X(x) = (1/2) f_X^{gG}(x; b_1, c_1, d_1) + (1/2) f_X^{gG}(x; b_2, c_2, d_2), \quad x \geq 0, \tag{26}$$

where f_X^{gG} is given by Eq. (12a) for $b_1 = 1.2$, $b_2 = 7$, $c_1 = c_2 = 0.296$, $d_1 = d_2 = 3$.

The pdf reconstruction procedure will be based *exclusively* on the numerical values of the $N = 16$ first moments $\mu_n^{(X)}$, as calculated by (trapezoidal) numerical integration (with some numerical errors) by considering 2000 equally-spaced points in the interval $[0, 5]$. The values of these moments are shown in Table 2. The error of each moment is identified and measured by means of the (*relative error*) ratio

$$\left| \frac{\mu_{n,exact}^{(X)} - \mu_n^{(X)}}{(1/2)\left(\mu_{n,exact}^{(X)} + \mu_n^{(X)}\right)} \right|, \quad \text{for each } n = 1, 2, \ldots, 16,$$

where $\mu_{n,exact}^{(X)} = (1/2)\sum_{i=1}^{2} c_i^{-(n/d_i)} \frac{\Gamma[(n/d_i)+b_i]}{\Gamma[b_i]}$ are the exact moments of (26). The results are presented in Fig. 3. Note that the relative error in low-order moments is less than the relative error in the high-order ones.

Now, the steps taken up towards the reconstruction are described in detail here below.

Step 0. *Numerical "admissibility" of the moment data*

The nonnegativity of the Hankel determinants (1a) and (1b) guarantees the *admissibility* of the (inexact) moment data $\mu_n^{(X)}$ (see [58] for more details). In the present case all Hankel determinants are positive, as expected. The values of the two last are $H_{16}^{(0)} = 6.7646409373 \times 10^{-4}$ and $H_{15}^{(1)} = 6.2891997484 \times 10^{-2}$.

Step 1. *Determination of the main-mass delimiter s*

The main-mass delimiter s is approximately determined by means of the first 14 moments as follows (see [42] for more details):

Let $x_{k,1}, x_{k,2}, \ldots, x_{k,k}$, *in ascending order*, be the roots of polynomials (see Section 2 for the notation)

$$P_k(x) := \left(H_{2k}^{(0)} H_{2k-2}^{(0)}\right)^{-(1/2)} D_k(x), \quad D_k(x) := \det \begin{bmatrix} \mu_0 & \mu_1 & \cdots & \mu_k \\ \vdots & \vdots & & \vdots \\ \mu_{k-1} & \mu_k & \cdots & \mu_{2k-1} \\ 1 & x & \cdots & x^k \end{bmatrix}, k = 7.$$

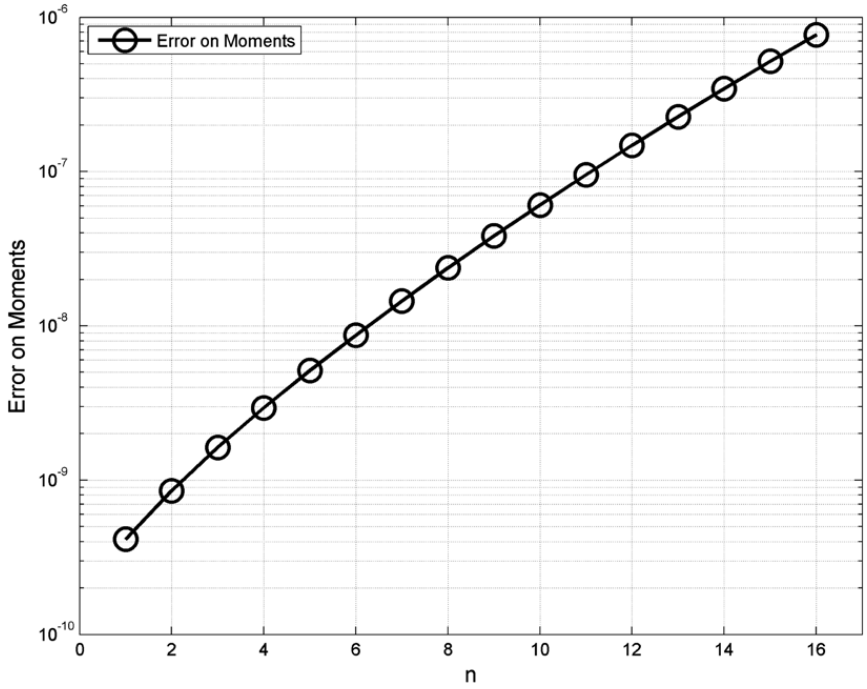

Fig. 3. The ratio $\left|\frac{\mu_{n,exact}^{(X)}-\mu_n^{(X)}}{(1/2)\left(\mu_{n,exact}^{(X)}+\mu_n^{(X)}\right)}\right|$ (error on moments), for $n = 1, 2, \ldots, 16$.

Table 3
The values of the lower and upper bounds ($L_{k,k}$ and $U_{k,k}$) calculated at the last root of the kth-order polynomial $P_k(x)$, $k = 3, 4, 5, 6, 7$.

$n = 2k$	$x_{k,k}$	$L_{k,k}$	$U_{k,k}$
6	$x_{3,3} = 3.1491155528$	0.6967445962	0.9999993576
8	$x_{4,4} = 3.3876509471$	0.8804232721	0.9999993576
10	$x_{5,5} = 3.5786243485$	0.9577713180	0.9999993589
12	$x_{6,6} = 3.7519683210$	0.9871658792	0.9999993921
14	$x_{7,7} = 3.9009760271$	0.9962099265	0.9999975752

Then, in accordance with a variant of Chebyshev–Stieltjes–Markov inequality [2,47], for the last root $x_{k,k}$, we have

$$L_{k,k} \equiv 1 - \lambda_k\left(x_{k,k}\right) \leq F\left(x_{k,k}\right) \leq \sum_{j=1}^{k} \lambda_k\left(x_{k,j}\right) \equiv U_{k,k}, \quad k = 7,$$

where $\lambda_k(x) := \left[\sum_{n=0}^{k} |P_n(x)|^2\right]^{-1}$ are the *Christoffel functions* [62].

The last root $x_{7,7}$ gives a very good estimate of the main-mass delimiter. In Table 3 the last roots $x_{k,k}$ of the polynomials $P_k(x)$ and the values of the lower and upper bounds, $L_{k,k}$ and $U_{k,k}$ are shown, for $2k = 6, 8, 10, 12, 14$. For example, the interval $\left[0, x_{7,7} = 3.9009760271\right]$, contains at least 99.6% of the total probability mass providing us with a very good estimate for the location of the main-mass region, and defining the semi-infinite interval $\left[x_{7,7}, +\infty\right)$ as the tail region. Note that even $x_{5,5}$ (corresponding to 10 moments) is a good indication for the main-mass delimiter. Let us note here that the roots of the polynomials are numerically realized by the MATLAB function *roots*.

Step 2. *Transformation to the normalized interval*

Under the knowledge of the main-mass delimiter s, the initial main-mass interval $A_{main}^{(X)} = [0, s = 3.9009760271]$ is transformed into the normalized interval $A_{main}^{(Y)} = [0, l = 0.8]$.

Step 3. *Transformation of moments* $\mu_n^{(X)}$ *to* $\mu_n^{(Y)}$

Following (21b), the initial moments $\mu_n^{(X)}$, $n = 1, \ldots, 15$, are transformed to $\mu_n^{(Y)}$, $n = 1, \ldots, 15$ (see Table 2). Comparing the values of the two moment datasets, it is clear that the transformed moments $\mu_n^{(Y)}$ exhibit a significantly smaller variation in their order of magnitude than the initial ones $\mu_n^{(X)}$. More precisely, the values of $\mu_n^{(X)}$ span 7 orders of magnitude (10^0–10^7) while $\mu_n^{(Y)}$ span 4 orders of magnitude (10^0–10^{-4}).

Table 4
The $(n-1)$-th order approximations $\tilde{d}_n\,(n-1)$, for $n = 10, \ldots, 14$ (Eq. (11)), by applying the extrapolation scheme to sequence $1/d_n$.

	$n=10$	$n=11$	$n=12$	$n=13$	$n=14$
$\tilde{d}_n\,(n-1)$	3.3528917105	3.3558414518	3.2594675606	3.1504291141	3.0784839727

Table 5
The location parameters $\{y_i\}_{i=1}^{I}$ of each KDF and the corresponding probability weights $\{p_i\}_{i=1}^{I}$, for $I = 11, 13, 15$.

I	i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
11	$y_i \cdot 10^2$	3.6	10	18	25	32	40	47	54	61	69	76				
	$p_i \cdot 10^2$	0	0	8.7	14	17	7.5	2.8	25	21	1.3	0				
13	$y_i \cdot 10^2$	3	9.2	15	21	27	33	40	46	52	58	64	70	76		
	$p_i \cdot 10^2$	0	0	2.1	14	6.5	21	0	6.1	11	27	9	0.3	0.0003		
15	$y_i \cdot 10^2$	2.6	8	13	18	24	29	34	40	45	50	56	61	66	72	77
	$p_i \cdot 10^2$	0	0	0.1	9.6	5.5	18	4.7	10	0	7.6	21	17	3.4	0.09	0.0001

Step 4. *Determination of the tail-form parameter d*

Now, the extrapolation scheme (11) is applied to the transformed moments $\mu_n^{(Y)}$ and the $(n-1)$-th order approximations $\tilde{d}_n\,(n-1)$ of d, for $n = 10, \ldots, 14$ are obtained (see Table 4). For example, the critical exponent d, as recovered by the 14 moments, is found to be $d = 3.0784839727$. Recall that the correct value is 3.00.

Step 5. *Choice of the index of complexity I, locations* $\{y_i\}_{i=1}^{I}$ *and bandwidth h*

If no other information is available, the suggested range of values for the index of complexity I is 10–15. It has been found, by means of various numerical experiments, that taking $I \leq 15$ resolves well unimodal and bimodal pdfs. Thus, in the present example we choose $I = 11, 13, 15$. A systematic procedure leading to a rational choice of I will be described in the next section.

Once the index of complexity I is chosen, the normalized main-mass interval $A_{main}^{(Y)} = [0, l]$ $(l \leq 1)$ is subdivided into equal subintervals by means of the partition $0 \equiv y_0^* < y_1^* < \cdots < y_{I-1}^* < y_I^* \equiv l$. On each subinterval, the middle points $y_i = y_{i-1}^* + (\Delta y^*/2)$, $\Delta y^* = y_i - y_{i-1}$, $i = 1, \ldots, I$, are chosen to serve as the location parameters of each KDF. Appropriate choices for the bandwidth parameter span the range $h = 0.06l$–$0.08l$, dependent on the choice of I. A sensitivity analysis concerning the values h/l is presented in Appendix B. In this example, the choice $h = 0.075l$ has been made. Note that we obtain practically identical reconstructions for all values $h = [0.06l$–$0.08l]$.

Step 6. *Numerical determination of the KDR approximants*

Once the parameters I, h and d are determined, for each i, $1 \leq i \leq I$, the generalized Gamma KDFs K^G are defined (see (A.1) of Appendix A) and the moments of the KDFs $B_{n,i}$, $i = 1, 2, \ldots, I$, $n = 1, \ldots, 15$ (see Eqs. (22b) and (A.5)) are numerically calculated. Then, by solving the constrained minimization problem (23) by means of the routine *lsqlin* of MATLAB we obtain the probability weights p_i, $i = 1, 2, \ldots, I$ (see Table 5). Accordingly, the KDR (Eq. (14a)) is determined and provides us with the approximate solution $f_Y^{(I)}(y)$, from which the target pdf $f_X(x)$ is approximated by means of the transformation $f_X^{(I)}(x) = (l/s)\, f_Y^{(I)}(lx/s)$.

Step 7. *Numerical convergence and comparison of the KDR approximants with the target pdf*

The numerical efficiency of the proposed algorithm is illustrated in Fig. 4, where the reconstructed $f_Y^{(I)}$, for $I = 11, 13, 15$, have been plotted, along with the transformed target pdf $f_Y(y)$ (see Eq. (21a) for the transformation). In a special zoom window, in the same figure, the pdfs' tail behavior is indicated on a logarithmic scale. The y-axis of the zoom window is $[0.8, 1.4]$ while the corresponding values of the $f_Y(y)$ are in the range $\left[10^{-2}, 10^{-20}\right]$. From this figure it is clearly seen that the reconstructed pdfs $f_Y^{(I)}$ are "uniformly valid", with both in the main-mass interval, as well as in the tail region. Approximants $f_Y^{(11)}$ and $f_Y^{(13)}$ have a very small discrepancy around the peaks, while $f_Y^{(15)}$ gives an almost perfect reconstruction along the whole range of the y-axis. In this (validation) example all three distances (24a), (24b) and (25) are used to measure the "goodness" of solution and the numerical convergence (see Table 6 for further indications).

Remark 5. As can be seen from Table 5, the numerical solution of problem (23) results in some "significant" weights p_i, while a few others have negligible contribution (i.e. $p_i \approx 0$). For example, for $I = 11$, the non-zero p_i come from 8 location parameters y_i, while for $I = 15$ the corresponding y_i are 12. Furthermore, numerical experience has shown that the solution remains essentially the same if we eliminate the KDFs with very small p_i values. These remarks have motivated an adaptive procedure permitting us to implement the following basic rule: *less KDFs must be used in regions of low probability and more KDFs in regions of high probability*.

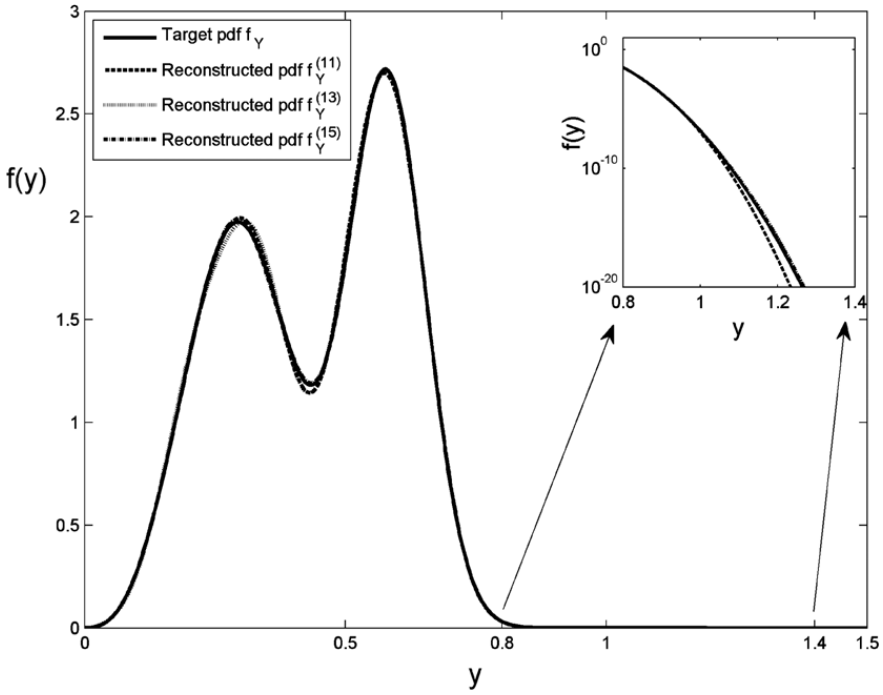

Fig. 4. Reconstruction of a bimodal pdf using 15 moments, for different values of the index of complexity I ($h = 0.075l$, $l = 0.8$).

Table 6
The *pdf-based* distances (24) and the *moment-based* semi-distance (25), for $N = 15$ given moments.

	$I = 11$	$I = 12$	$I = 13$	$I = 14$	$I = 15$
$d_{\max}\left(f_Y, f_Y^{(I)}\right) \cdot 10^2$	3.3180261581	7.1775754147	6.6115041152	4.2926330115	1.5895692705
$d_1\left(f_Y, f_Y^{(I)}\right) \cdot 10^2$	0.7176615729	1.6566607110	1.3323740431	0.9327570856	0.3215936680
$d_{mom}\left(\left\{\mu_n^{(Y)}\right\}, \left\{\hat{\mu}_n^{(Y)}\right\}\right) \cdot 10^6$	1.5593729896	0.1692662859	0.0930681286	0.0044908229	0.0004312789

6. Algorithm for optimal selection of KDFs' location: a sequentially adaptive algorithm (SAA)

A possible drawback of the previously presented numerical solution method is that the index of complexity I is given an arbitrary value (with no rule) and the location parameters $\{y_i\}_{i=1}^I$ are considered to be uniformly distributed, without considering any "optimality" criterion. Here, we propose an iterative, adaptive algorithm, having similarities with updating and learning in *probabilistic expert systems* [63,64] and with training algorithms in *neural networks* [65]. The philosophy behind this scheme is the *principle of parsimony*, i.e. *to use the minimal number of parameters, under the condition that the prediction power of available data has been fully exploited*.

6.1. Description of the iterative, adaptive algorithm

The proposed algorithm, called *S*equentially *A*daptation *A*lgorithm (SAA), is depicted in a number of steps, in Fig. 5. According to this scheme, the *"basis" system* $\left\{I, \{y_i\}_{i=1}^I\right\}$ "propagates" using a *death/birth procedure*, based on the current values of $\{p_i\}_{i=1}^I$, in conjunction with the following set of criteria (*move types*) (see steps 5 and 6 in Fig. 5):

(a) the death of insignificantly weighted KDFs; see step 5.1,
(b) the birth of new y_i's around significantly weighted KDFs; see step 6.1,
(c) splitting one KDF component into two; see steps 5.2, 6.1,
(d) merging two KDF components into one; see steps 5.3, 6.2.

The move (a) saves computational time simplifying the KDR by allowing to get rid of KDFs with insignificant weights. The moves (b) and (c) replace KDFs of high probability weight with two nearly located KDFs. The move type (d) merges two adjacent components when their location distance is relatively small. Let us note that the moves (a) and (d) permit us to avoid the progressive "impoverishment" or "degeneracy" [66] of the resulting KDR model. At the end, after applying the moves (a)–(d) to a specific basis system, a new "refreshing" one is produced, for which the corresponding probability weights are calculated by means of (23).

Sequentially Adaptive Algorithm (SAA)		
Step 0.	*Initialization*. For $k=0$, set I_0 (a small number, e.g., 2-6); $I_{\max}$ (e.g., 12); $k_{\max}$ (e.g., 10-15); The location parameters $\{y_{i,0}\}_{i=1}^{I_0}$, with $y_{1,0}<\ldots<y_{I_0,0}$ (e.g.,uniformly distributed); a *death threshold* ε_{death} (e.g., 10^{-10}); an *accuracy threshold* ε_{accur} (e.g., 10^{-6}); a *nearness threshold* ε_{near} (e.g., 10^{-2}).	
For any k,		
Step 1.	Calculate the probability weights $\{p_{i,k}\}_{i=1}^{I_k}$, using the currently available basis system $\left\{I_k,\{y_{i,k}\}_{i=1}^{I_k}\right\}$ and evaluate the approximate pdf $f_{Y,k}^{(I_k)}(y)$, and its moments $\hat{\mu}_{n,k}$.	
Step 2.	*Stopping Criteria.* If $d_{mom,k}\left(\left\{\mu_n^{(Y)}\right\}_{n=1}^{N},\left\{\hat{\mu}_{n,k}^{(Y)}\right\}_{n=1}^{N}\right)<\varepsilon_{accur}$, stop. Otherwise goto Step 3.	
Step 3.	*Location of maximum probability weight.* For $1\le i\le I_k$, find the index $i_{\max}$ corresponding to the maximum probability weight among all weights, i.e. $i_{\max}:\ p_{i_{\max},k}=\max_{1\le i\le I_k} p_{i,k}$.	
Step 4.	*Closeness.* For $1\le i\le I_k$, find all indices i_{near} such that $\left\lvert y_{i_{near},k}-y_{i_{near}+1,k}\right\rvert<\varepsilon_{near}$.	
Step 5.	***Deaths.***	
	5.1	*Zero weight.* For $1\le i\le I_k$, remove locations $y_{i,k}$ when the corresponding $p_{i,k}<\varepsilon_{death}$.
	5.2	*Maximum weight.* Remove location $y_{i_{\max,k}}$.
	5.3	*Close locations.* Remove locations $y_{i_{near},k}$ and $y_{i_{near}+1,k}$.
Step 6.	***Births.***	
	6.1	*Maximum weight.* For $i=i_{\max}$, define the pair $\left\{y_k^{(1)},y_k^{(2)}\right\}=\left\{y_{i_{\max},k}-r\Delta y_{i_{\max,k}},\ y_{i_{\max},k}+r\Delta y_{i_{\max,k}}\right\}$ (usually $r=1/4$ or $1/3$), a) If $y_k^{(1)}>0$ and $y_k^{(2)}<l$, then *add* the pair $\left\{y_k^{(1)},y_k^{(2)}\right\}$ as location parameters, b) If $y_k^{(1)}<0$, then *add only* the $y_k^{(2)}$ to the location parameter set, c) If $y_k^{(2)}>l$, then *add only* the $y_k^{(1)}$ to the location parameter set.
	6.2	*Close locations.* For $i=i_{near}$, *add* the location $y_{i_{mid},k}=\dfrac{y_{i_{near},k}+y_{i_{near}+1,k}}{2}$.
Step 7.	Sort all active (alive) locations and obtain the new basis system $\left\{I_{k+1},\{y_{i,k+1}\}_{i=1}^{I_{k+1}}\right\}$.	
Step 8.	*Loop.* If $I_{k+1}\le I_{\max}$ and $k<k_{\max}$ then set $k=k+1$ and return to Step 1. Otherwise stop.	

Fig. 5. Basic steps of the SAA.

Stopping criteria. The regular stopping criterion is formulated in terms of the moment-based distance (25), i.e. when $d_{mom}\left(\{\mu_n\}_{n=1}^{N},\{\hat{\mu}_n\}_{n=1}^{N}\right)$ is lower than a given threshold. If this criterion fails, the procedure will be stopped when either the number of iterations reaches a maximum value $k_{\max}$, or the index of complexity I exceeds a prespecified value $I_{\max}$.

Example 3 (*Application of the SAA to* Example 2)**.** Recall that

- the *input data* $\left\{\mu_n^{(X)},N\right\}$ are summarized by Table 2,
- the *calculated intermediate data* $\left\{\mu_n^{(Y)},s,d\right\}$ are given in Tables 2–4,
- the *user-defined numerical parameters* are $h=0.075l$, $l=0.8$.

As the *initial basis system* we consider the following two-point set $\left\{I_0 = 2, \left\{y_{1,0} = 0.05, y_{2,0} = 0.7\right\}\right\}$. The SAA is applied to the above data, producing the *output data* $\left\{I_k, \left\{y_{i,k}\right\}_{i=1}^{I_k}, \left\{p_{i,k}\right\}_{i=1}^{I_k}\right\}$, for $k = 0, 1, \ldots, 17$, and a sequence of approximants $f_{Y,k}^{(I_k)}(y)$, $k = 0, 1, \ldots, 17$. The propagation of the iteration algorithm, shown in Fig. 6, can be subdivided into three phases:

- *Initialization phase: the zero iteration*($k = 0$, Fig. 6(a)). The approximant $f_{Y,0}^{(2)}(y)$ shares *only* the same tail behavior with the target pdf (see the special zoom window in the same figure).
- *Evolution phase: the first iterations* ($k = 1$–7, Fig. 6(b)–(f)). As the KDF locations' propagate, through iterations, the approximants adjust their shape in order to "capture" the main features of the target pdf in the main-mass interval, while the tail behavior remains the same.
- *Limit phase: the last iterations* ($k = 8$–17, Fig. 6(g)–(h)). The convergence has been essentially reached and thus the variation of the approximants is negligible, and the error is practically constant (see Fig. 6(i)). □

7. Discussion and conclusions

In this paper a stable and efficient numerical solution to the truncated Stieltjes moment problem, called KDEM, has been presented. A key feature of this method is the *a priori* exact characterizations of the data space and the solution space, enabling a Thikhonov (conditional) regularization of the inversion of the moment data. Another feature of the proposed solution method is the determination of the pdf's tail behavior directly, through moments. Let it be noted that even though, theoretically, the pdf tail behavior is defined in terms of high-order moment asymptotics, the sought-for asymptotic limits are numerically realized on the basis of a small number of moments, after applying the Richardson extrapolation (acceleration) scheme. More specifically, the tail parameters are efficiently numerically determined by using only 5–6 moments for unimodal cases (see Table 1 and Fig. 1) and 5–10 for bimodal cases (see Table 4). The KDEM has been tested in various cases (unimodal and bimodal target pdfs, inexact and noisy moment data), giving in almost all cases very satisfactory results.

Various important questions still remain to be addressed. Among them we refer to the following:

- Extension of the KDEM to the determination of pdfs supported on the whole real line (Hamburger moment problem), by simultaneous determination of the two tails.
- Extension of the KDEM to the bivariate and multivariate cases.
- Use of different "moment" information (fractional moments, see e.g., [67–69] with the KDR).
- Use of different methods (e.g. the maximum entropy method [23]) for the calculation of the probability weights p_i of the KDR.
- Use of the KDEM in stochastic systems (see e.g., [70–72]).

Acknowledgments

We are grateful to Professor Mario Bertero for his constructive criticism and for encouraging this work. We thank Professor Gabriele Inglese for insightful remarks, Professor Ward Whitt and Dr. David Lucantoni for provide us with their works and useful remarks. We thank the anonymous reviewer for his/her comments and suggestions.

Appendix A. The generalized Gamma kernel density function

In this work, the generalized Gamma (gG) KDF $K^{gG}(x)$ has been chosen to serve as an appropriate kernel, from the target space T (see Section 3.1). Other KDF choices (e.g. Gaussian, lognormal, etc.) are avoided as they do not comply with the boundary conditions of the solution space T.

The generalized Gamma KDF $K^{gG}(x)$ is defined to be the gG pdf $f_X^{gG}(x)$ under appropriate definition of the kernel parameters: *location parameter* w, *bandwidth parameter* h and *tail form parameter* d, i.e.

$$K^{gG}(x; w, h, d) \equiv f_X^{gG}(x; b, c, d) = \frac{dc^b}{\Gamma[b]} x^{bd-1} \exp\left\{-cx^d\right\}, \quad x \in [0, \infty), \tag{A.1}$$

where the (tail form) parameter d is calculated from the asymptotic behavior of the moment sequence, as explained in Section 3.2, and the parameters $b \equiv b(w, h; d)$ and $c \equiv c(w, h; d)$ are defined by solving the algebraic nonlinear system

$$w \equiv x_{mode}\left(f_X^{gG}(x; b, c, d)\right), \tag{A.2}$$

$$h \equiv \sigma\left(f_X^{gG}(x; b, c, d)\right). \tag{A.3}$$

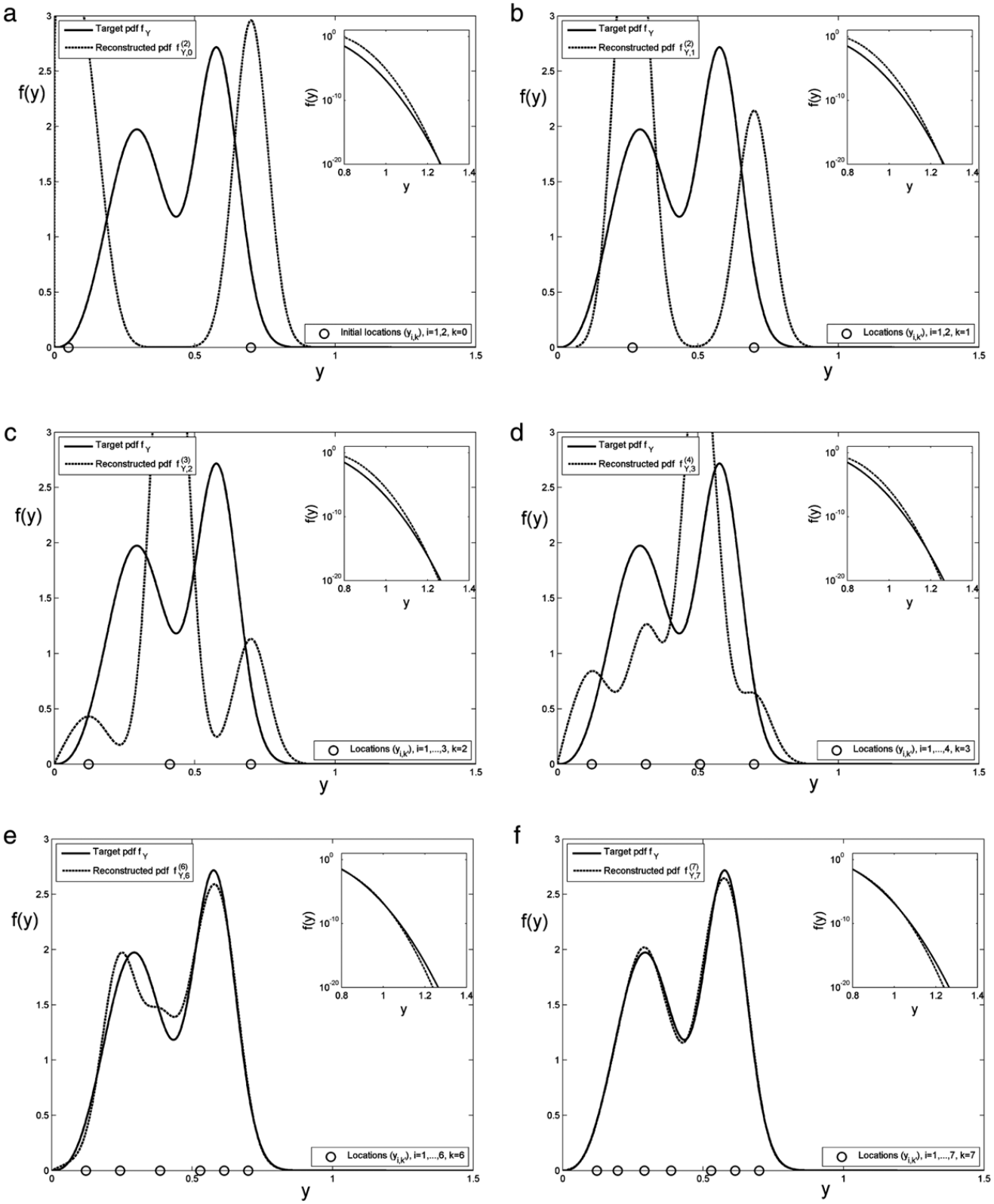

Fig. 6. In (a)–(h). The reconstructed $f_{Y,K}^{I_k}(y)$ are plotted, for $k = 0, 1, 2, 3, 6, 7, 8, 12$, by using the SAA, along with the target pdf (the same target pdf as in Fig. 3). In (i), the Mean Relative Error (25) is plotted for all k iterations.

That is

- the *location parameter* w is defined to be the mode (most probable value) of f_X^{gG};
- the *bandwidth parameter* h is defined to be the standard deviation of f_X^{gG}.

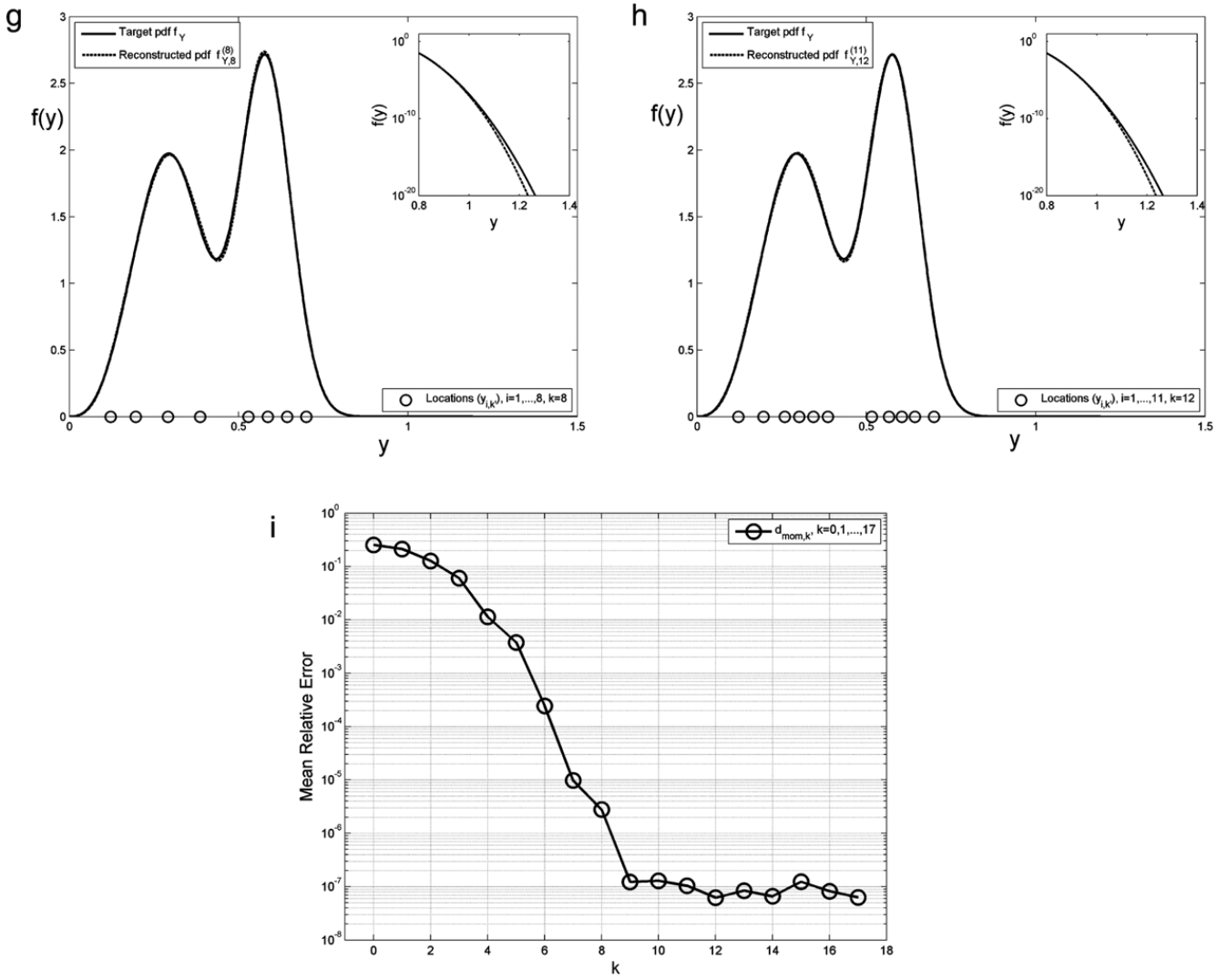

Fig. 6. (*continued*)

Recall that

$$x_{mode} : df_X^{gG}(x_{mode})/dx = 0 \Rightarrow x_{mode}\left(f_X^{gG}\right) = c^{-(1/d)}\left(b-\frac{1}{d}\right)^{(1/d)},$$

$$\sigma^2\left(f_X^{gG}\right) = c^{-(2/d)}\frac{\Gamma\left[b+\frac{2}{d}\right]}{\Gamma[b]} - \left[c^{-(1/d)}\frac{\Gamma\left[b+\frac{1}{d}\right]}{\Gamma[b]}\right]^2.$$

Using the above results, Eqs. (A.2) and (A.3) can be written in the form

$$w = c^{-(1/d)}\left(b-\frac{1}{d}\right)^{(1/d)}, \tag{A.4}$$

$$h^2 = c^{-(2/d)}\frac{\Gamma\left[b+\frac{2}{d}\right]}{\Gamma[b]} - \left[c^{-(1/d)}\frac{\Gamma\left[b+\frac{1}{d}\right]}{\Gamma[b]}\right]^2. \tag{A.5}$$

Expressing $c^{-(1/d)}$ in terms of w, b and d from Eq. (A.4), and substituting in Eq. (A.5), we obtain

$$\left(b-\frac{1}{d}\right)^{(2/d)} - \left(\frac{w}{h}\right)^2\left[\frac{\Gamma\left[b+\frac{2}{d}\right]}{\Gamma[b]} - \left(\frac{\Gamma\left[b+\frac{1}{d}\right]}{\Gamma[b]}\right)^2\right] = 0, \tag{A.6}$$

which provide us the value of $b \equiv b(w, h; d)$. Then, c is obtained by

$$c = w^{-d}\left(b-\frac{1}{d}\right). \tag{A.7}$$

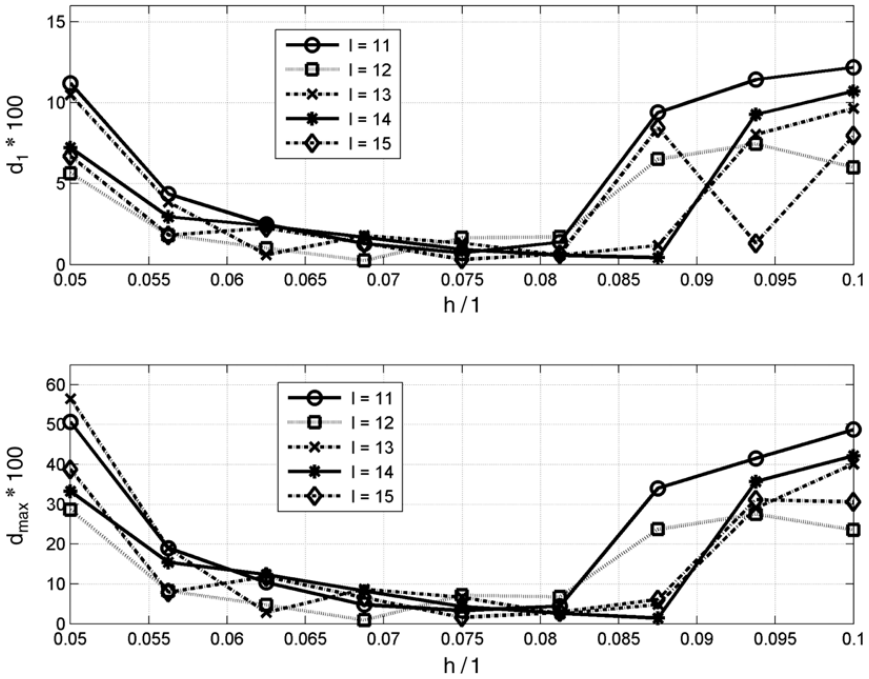

Fig. B.1. The distances $d_1\left(f_Y, f_Y^{(l)}\right)$ and $d_{\max}\left(f_Y, f_Y^{(l)}\right)$ as a function of h/l, $(l = 0.8)$, for different values of I and given number of moments N.

The moments of the gG kernel, located at any w, are given by

$$B_n(w) = \int_0^{+\infty} x^n K^{gG}(x; w, h, d)\ dx = c^{-(n/d)}\ \frac{\Gamma\left[\frac{n}{d} + b\right]}{\Gamma[b]}, \quad n = 1, 2, 3, \ldots, \tag{A.8}$$

where $b \equiv b(w, h; d)$ and $c \equiv c(w, h; d)$ are the solutions of system (A.6) and (A.7).

The above results generalize similar ones for the Gamma KDF $(d = 1)$ developed and used in [73].

Appendix B. Selection of the bandwidth *h* in the case of equally spaced partition

Our goal here is to justify the selection of the numerical parameter h (bandwidth of the KDF). For this purpose, a typical member of the target space (e.g., the density of the form (26)) is considered, and the two distances $d_{\max}\left(f_Y, f_Y^{(l)}\right)$ and $d_1\left(f_Y, f_Y^{(l)}\right)$ are plotted (see Fig. B.1), as a function of the ratio h/l $(l = 0.8)$ for different values of I and given number of moments $N = 16$.

As seen from the figure, both distances share a very similar pattern. In particular, as I increases, three error regions are clearly indicated:

- $h/l \in [0.05, 0.06]$; the error decreases,
- $h/l \in [0.06, 0.08]$; the error attains the lower values and remains practically constant,
- $h/l \in [0.08, 0.1]$; the error increases.

Based on these results, we propose as the "optimal" h range the interval $[0.06l, 0.08l]$, which ensures an "*almost*" uniformly valid reconstruction. Let it be noted that similar results have also been obtained for various other members of the target space T.

References

[1] T.J. Stieltjes, Recherchers sur les fractions continues, Annales de la Faculte des Sciences de Toulouse 8 (1884) 1–122.
[2] N.I. Akhiezer, The Classical Moment Problem, Oliver and Boyd, London, 1965.
[3] J.A. Shohat, J.D. Tamarkin, The problem of moments, American Mathematical Survey (1943).
[4] J.M. Stoyanov, Counterexamples in Probability, Wiley, New York, 1987.
[5] C. Berg, On powers of Stieltjes moment sequences, I, Journal of Theoretical Probability 18 (2005) 871–889.
[6] C. Berg, On powers of Stieltjes moment sequences, II, Journal of Computational and Applied Mathematics 199 (2007) 23–38.
[7] A. Gut, On the moment problem, Bernoulli 8 (2002) 407–421.
[8] G.D. Lin, On the moment problems, Statistics & Probability Letters 35 (1997) 85–90.
[9] A.G. Pakes, Structure of Stieltjes classes of moment-equivalent probability laws, Journal of Mathematical Analysis and Applications 326 (2007) 1268–1290.

[10] A.G. Pakes, W.-L. Hung, J.-W. Wu, Criteria for the unique determination of probability distributions by moments, Journal of the Australian Mathematical Society 43 (2001) 101–111.
[11] J.M. Stoyanov, Krein condition in probabilistic moment problems, Bernoulli 6 (2000) 939–949.
[12] D. Bertsimas, I. Popescu, On the relation between option and stock prices: a convex optimization approach, Operations Research 50 (2002) 358–374.
[13] D. Bertsimas, I. Popescu, Optimal inequalities in probability theory: a convex optimization approach, SIAM Journal on Optimization 15 (2005) 780–804.
[14] D. Bessis, C.R. Handy, Moment problem formulation of the Schrodinger equation, Numerical Algorithms 3 (1992) 1–16.
[15] C.R. Handy, D. Bessis, D. Morley, Generating quantum energy bounds by the moment method: a linear-programming approach, Physical Review A 15 (1988) 4557–4569.
[16] C.R. Handy, G.L. Ndow, Euclidean time formulation of the eigenvalue moment method: a moment problem-convexity analysis of Barnsley's theorem, Journal of Physics A: Mathematical and General 25 (1992) 2669–2681.
[17] J.B. Lasserre, Bounds on measures satisfying moment conditions, The Annals of Applied Probability 12 (2002) 1114–1137.
[18] J.B. Lasserre, T. Prieto-Rumeau, M. Zervos, Pricing a class of exotic options via moments and SDP relaxations, Mathematical Finance 16 (2006) 469–494.
[19] P. Weidelt, The relationship between the spectral function and the underlying conductivity structure in 1-D magnetotellurics, Geophysical Journal International 161 (2005) 566–590.
[20] D. Poland, Time evolution of polymer distribution functions from moment equations and maximum-entropy methods, The Journal of Chemical Physics 111 (1999) 8214–8224.
[21] A. Zarzo, J.C. Angulo, J. Antolin, R.J. Yanez, Maximum-entropy analysis of one-particle densities in atoms, Zeitschrift fur Physik D Atoms, Molecules and Clusters 37 (1996) 295–299.
[22] A. Berera, P.A. Martin, Inverting the Sachs–Wolfe formula: an inverse problem arising in early-universe cosmology, Inverse Problems 15 (1999) 1393–1404.
[23] U. Alibrandi, G. Ricciardi, Efficient evaluation of the pdf of a random variable through the kernel density maximum entropy approach, International Journal for Numerical Methods in Engineering 75 (2008) 1511–1548.
[24] K. Sobczyk, Stochastic Differential Equations with Applications to Physics and Engineering, Kluwer Academic Publishers, Boston, 1991.
[25] K. Sobczyk, J. Trebicki, Maximum entropy principle and nonlinear stochastic oscillators, Physica A: Statistical and Theoretical Physics 193 (1993) 448–468.
[26] J.R. Klauder, K.A. Penson, J.-M. Sixdeniers, Constructing coherent states through solutions of Stieltjes and Hausdorff moment problems, Physical Review A 64 (2001) 013817.
[27] K.A. Penson, A.I. Solomon, New generalized coherent states, Journal of Mathematical Physics 40 (1999) 2354–2363.
[28] K.A. Penson, P. Blasiak, G. Duchamp, A. Horzela, A.I. Solomon, Hierarchical Dobinski-type relations via substitution and the moment problem, Journal of Physics A: Mathematical and General 37 (2004) 3475–3487.
[29] R.B. Diemer, J.H. Olson, A moment methodology for coagulation and breakage problems: Part 2—moment models and distribution reconstruction, Chemical Engineering Science 57 (2002) 2211–2228.
[30] V. John, I. Angelov, A.A. Öncül, D. Thévenin, Techniques for the reconstruction of a distribution from a finite number of its moments, Chemical Engineering Science 62 (2007) 2890–2901.
[31] R. McGraw, S. Nemesure, S.E. Schwarts, Properties and evolution of aerosols with size distributions having identical moments, Journal of Aerosol Science 29 (1998) 761–772.
[32] P. Buchen, M. Kelly, The maximum entropy distribution of an asset inferred from option prices, Journal of Financial and Quantitative Analysis 31 (1996) 143–159.
[33] S.A. El-Wakil, E.M. Abulwafa, M.A. Abdou, A. Elhanbaly, Maximum-entropy approach with higher moments for solving Fokker–Planck equation, Physica A 315 (2002) 480–492.
[34] H. Geman, M. Yor, Bessel processes, Asian options and perpetuities, Mathematical Finance 3 (1993) 349–375.
[35] X. Wu, Calculation of maximum entropy densities with application to income distribution, Journal of Econometrics 115 (2003) 347–354.
[36] M. Bertero, C. De Mol, E.R. Pike, Linear inverse problems with discrete data. I: General formulation and singular system analysis, Inverse Problems 1 (1985) 301–330.
[37] G. Talenti, Recovering a function from a finite number of moments, Inverse Problems 3 (1987) 501–517.
[38] G.A. Athanassoulis, P.N. Gavriliadis, The Hausdorff moment problem solved by using kernel density functions, Probabilistic Engineering Mechanics 17 (2002) 273–291.
[39] A.N. Tikhonov, V.Y. Arsenin, Solutions of Ill-Posed Problems, Wiley, New York, 1977.
[40] A.N. Tikhonov, A.V. Goncharsky, Ill-Posed Problems in the Natural Sciences, Mir Pub., Moscow, 1987.
[41] P.N. Gavriliadis, Moment information for probability distributions, without solving the moment problem, I: where is the mode? Communications in Statistics. Theory and Methods 37 (2008) 671–681.
[42] P.N. Gavriliadis, G.A. Athanassoulis, Moment information for probability distributions, without solving the moment problem. II: Main-mass, tails and shape approximation, Journal of Computational and Applied Mathematics 229 (2009) 7–15.
[43] K. Bandyopadhyay, K. Bhattacharyya, A.K. Bhattacharyya, On some problems of the maximum entropy ansatz, Pranama Journal of Physics 54 (2000) 365–375.
[44] K. Bandyopadhyay, K. Bhattacharyya, A.K. Bhattacharyya, Maximum-entropy principle with moment recursion relations as constraints, Physical Review A 63 (2001) 064101–064104.
[45] J.C. Bronski, D.W. McLaughlin, Rigorous estimates of the tails of the probability distribution function for the random linear shear model, Journal of Statistical Physics 98 (2000) 897–915.
[46] E.B. Cobb, B. Haris, The characterization of the solution sets for the generalized reduced problems and its application to numerical integration, Siam Review 8 (1966) 86–99.
[47] M.G. Krein, A.A. Nudelman, The Markov Moment Problem and Extremal Problems, AMS, Providence RI, 1977.
[48] W.D. Smith, The Gamma function revisited, unpublished manuscript, 2006. http://scorevoting.net/WarrenSmithPages/homepage/gammprox.pdf.
[49] J. Abate, G.L. Choudhury, D.M. Lucantoni, W. Whitt, Asymptotic analysis of tail probabilities based on the computation of moments, Annals of Applied Probability 5 (1995) 983–1007.
[50] C.M. Bender, S.A. Orszag, Advanced Mathematical Methods for Scientists and Engineers, McGraw-Hill, New York, 1978.
[51] S. Blinnikov, R. Moessner, Expansions for nearly Gaussian distributions, Astronomy and Astrophysics Supplement Series 130 (1998) 193–205.
[52] M.J.D. Powell, Approximation Theory and Methods, Cambridge University Press, 1981.
[53] W.B. Jurkat, G.G. Lorentz, Uniform approximation polynomials with positive coefficients, Duke Mathematical Journal 28 (1961) 463–473.
[54] G.K. Er, A method for multi-parameter pdf estimation of random variables, Structural Safety 20 (1998) 25–36.
[55] M. Frontini, A. Tagliani, Entropy-convergence in Stieltjes and Hamburger moment problem, Applied Mathematics and Computation 88 (1997) 39–51.
[56] A. Tagliani, On the application of maximum entropy to the moments problem, Journal of Mathematical Physics 34 (1993) 326–337.
[57] Y. Katznelson, An Introduction to Harmonic Analysis, Wiley, New York, 1968.
[58] P.N. Gavriliadis, G.A. Athanassoulis, Moment data can be analytically completed, Probabilistic Engineering Mechanics 18 (2003) 329–338.
[59] X. Li, J. Le, P. Gopalakrishnan, L. Pileggi, Asymptotic probability extraction for nonnormal performance distributions, IEEE TCAD 26 (2007) 16–37.
[60] T.F. Coleman, Y. Li, A reflective Newton method for minimizing a quadratic function subject to bounds on some of the variables, SIAM Journal on Optimization 6 (1996) 1040–1058.
[61] P.E. Gill, W. Murray, M.H. Wright, Practical Optimization, Academic Press, London, 1981.
[62] P. Nevai, Geza Freud, Orthogonal polynomials and Christoffel functions. A case study, Journal of Approximation Theory 48 (1986) 3–167.
[63] J.S. Liu, R. Chen, Sequential Monte-Carlo methods for dynamic systems, Journal of American Statistical Association 93 (1998) 1032–1044.

[64] S. Richardson, P. Green, On Bayesian analysis of mixtures with unknown number of components, Journal of the Royal Statistical Society B 59 (1997) 731–792.
[65] A. Alexandridis, H. Sarimveis, G. Bafas, A new algorithm for online structure and parameter adaptation of RBF networks, Neural Networks 16 (2003) 1003–1017.
[66] N. Chopin, A sequential particle filter method for static models, Biometrika 89 (2002) 539–552.
[67] P.N. Inverardi, A. Petri, G. Pontuale, A. Tagliani, Stieltjes moment problem via fractional moments, Applied Mathematics and Computation 166 (2005) 664–677.
[68] G. Cottone, M. Di Paola, On the use of fractional calculus for the probabilistic characterization of random variables, Probabilistic Engineering Mechanics 24 (2009) 321–330.
[69] H. Gzyl, A. Tagliani, Stieltjes moment problem and fractional moments, Applied Mathematics and Computation 216 (2010) 3307–3318.
[70] C. Pantano, B. Shotorban, A least-squares dynamic approximation method for evolution of uncertainty in initial conditions of dynamical systems, Physical Review E 76 (2007) 066705–066713.
[71] T.P. Sapsis, G.A. Athanassoulis, New partial differential equations governing the response-excitation joint probability distributions of nonlinear systems under general stochastic excitation, Probabilistic Engineering Mechanics 23 (2008) 289–306.
[72] B. Shotorban, Dynamic least-squares kernel density modeling of Fokker–Planck equations with application to neural population, Physical Review E 81 (2010) 046706.
[73] G.A. Athanassoulis, K.A. Belibassakis, Probabilistic description of metocean parameters by means of kernel density models 1. Theoretical background and first results, Applied Ocean Research 24 (2002) 1–20.