Abstract

Hausdorff moment problem in the approach of maximum entropy is reconsidered. Some aspects, as existence condition, entropy convergence, entropy decreasing, stability are studied through a unified approach. Results known in literature and some new ones are illustrated © 1999 Elsevier Science Inc. All rights reserved.

Keywords: Convex hull; Entropy; Hankel determinant; Moment problem

1. Introduction

The recovering of a probability density \( f(x) \) of which a finite number of moments are known represents a classical indetermined problem as there are, in general, an infinite variety of functions with the first \( M + 1 \) moments \( \mu_0, \ldots, \mu_M \). To stabilize the problem it is necessary to impose some further condition which will act as a constraint within the function space and lead to a unique solution. The maximum entropy (ME) approach [1] offers a definite procedure for the reconstruction of an approximant \( f_M(x) \) under the condition that the first \( M + 1 \) moments be equal to the true moments \( \mu_j, j = 0, \ldots, M \). In particular we study Hausdorff finite moment problem in which the goal is to recover a probability density \( f(x) : [0, 1] \to \mathbb{R}_+ \) from the knowledge of finitely many moments \( \mu_j, j = 0, \ldots, M \).
Several interesting analytical and computational aspects of the Hausdorff moment problem have been tackled in literature in the framework of the ME, as (1) existence of the solution [2], (2) convergence [3–5], (3) entropy decreasing, (4) stability [6]. All the quoted aspects have been tackled by resorting to spread of techniques. In this paper we reconsider some analytical and computational aspects of the problem by a unified approach allowing us to obtain both results previously found and new ones concerning the stability.

2. Background

2.1. ME formulation

Defined the (Shannon) entropy $H[f]$ of the probability density $f(x)$

$$H[f] = - \int_0^1 f(x) \ln f(x) \, dx$$

and the corresponding moments

$$\int_0^1 x^j f(x) \, dx = \mu_j \quad j = 0, \ldots, M, \quad \mu_0 = 1,$$

the ME approximant is obtained by maximizing Eq. (1) constrained by Eq. (2). By introducing the Lagrangian with the multipliers $\lambda_0, \ldots, \lambda_M$

$$L = - \int_0^1 f(x) \ln f(x) \, dx - (\lambda_0 - 1) \left( \int_0^1 f(x) \, dx - \mu_0 \right)$$

$$- \sum_{j=1}^M \lambda_j \left( \int_0^1 x^j f(x) \, dx - \mu_j \right),$$

one obtains by standard procedure the approximant [7]

$$f_M(x) = \exp \left( - \sum_{j=0}^M \lambda_j x^j \right),$$

to be supplemented by the condition that the first $M + 1$ moments be given by $\mu_j$, $j = 0, \ldots, M$

$$\int_0^1 x^j f_M(x) \, dx = \mu_j \quad j = 0, \ldots, M.$$

The uniqueness of the ME solution, if it exists, is guaranteed by convexity of the functional $\Gamma(\lambda_1, \ldots, \lambda_M)$ defined by [7]
\[ \Gamma(\lambda_1, \ldots, \lambda_M) = \sum_{j=1}^{M} \lambda_j \mu_j + \ln \left( \int_0^1 \exp \left( - \sum_{j=1}^{M} \lambda_j x^j \right) \, dx \right) \] (6)

The minimization of Eq. (6) provides a practical tool to calculate \( \lambda_1, \ldots, \lambda_M \), while \( \lambda_0 \) is obtained from first equation of Eq. (5).

If \( \mu_0, \ldots, \mu_M \) are assigned, by integrating by parts Eq. (5) allows us to obtain next unknown moments \( \mu_j, j > M \) as function of the first \( M + 1 \) and of \( \lambda_j, j = 0, \ldots, M \)

\[ (n + 1) \mu_n - 1 + \sum_{j=1}^{M} j \lambda_j (\mu_j - \mu_{j+n}) = 0 \quad n = 1, 2, \ldots \] (7)

2.2. The moment space and Hankel’s determinants

Let \( D^M \subset \mathbb{R}_+ \) be the convex hull of the curve \( \{x, x^2, \ldots, x^M\}, x \in [0, 1] \). \( D^M \) is called in literature \( M \)-moment space and is known to be a convex body, i.e. its interior (int \( D^M \)) includes a ball of dimension \( M \) [8]. It is known that if (a) the point \( \mu = (\mu_1, \ldots, \mu_M) \) is outside \( D^M \), the corresponding finite Hausdorff moment problem does not admit any solution, (b) \( \mu \in \partial D^M \) (\( \partial D^M \) is the boundary of \( D^M \)) the only distribution having \( (\mu_1, \ldots, \mu_M) \) as first moments is a (uniquely determined) convex combination of Dirac’s delta, (c) \( \mu \) belongs to int \( D^M \) many infinitely distributions exist.

The previous conditions about existence conditions can be algebraically obtained by introducing the Hankel determinants \( \Delta l_0, \Delta u_0, \Delta l_1, \Delta l_1, \ldots, \Delta l_M, \Delta l_M \), where \( M = 2N \)

\[ \Delta l_0 = \mu_0, \quad \Delta l_2 = \begin{vmatrix} \mu_0 & \mu_1 \\ \mu_1 & \mu_2 \end{vmatrix}, \quad \Delta l_4 = \begin{vmatrix} \mu_0 & \mu_1 & \mu_2 \\ \mu_1 & \mu_2 & \mu_3 \\ \mu_2 & \mu_3 & \mu_4 \end{vmatrix}, \]

\[ \cdots, \quad \Delta l_{2N} = \begin{vmatrix} \mu_0 & \cdots & \mu_N \\ \vdots & \vdots & \vdots \\ \mu_N & \cdots & \mu_{2N} \end{vmatrix}, \]

\[ \Delta u_2 = \mu_1 - \mu_2, \quad \Delta u_4 = \begin{vmatrix} \mu_1 - \mu_2 & \mu_2 - \mu_3 \\ \mu_2 - \mu_3 & \mu_3 - \mu_4 \end{vmatrix}. \] (8)
\[ \Delta u_6 = \begin{vmatrix} \mu_1 - \mu_2 & \mu_2 - \mu_3 & \mu_3 - \mu_4 \\ \mu_2 - \mu_3 & \mu_3 - \mu_4 & \mu_4 - \mu_5 \\ \mu_3 - \mu_4 & \mu_4 - \mu_5 & \mu_5 - \mu_6 \end{vmatrix} \]

\[ \cdots, \Delta u_{2N} = \begin{vmatrix} \mu_1 - \mu_2 & \cdots & \mu_N - \mu_{N+1} \\ \vdots & \vdots & \vdots \\ \mu_N - \mu_{N+1} & \cdots & \mu_{2N-1} - \mu_{2N} \end{vmatrix} , \quad (9) \]

\[ M = 2N + 1 \]

\[ \Delta l_1 = \mu_1, \quad \Delta l_3 = \begin{vmatrix} \mu_1 & \mu_2 \\ \mu_2 & \mu_3 \end{vmatrix} , \quad \Delta l_5 = \begin{vmatrix} \mu_1 & \mu_2 & \mu_3 \\ \mu_2 & \mu_3 & \mu_4 \\ \mu_3 & \mu_4 & \mu_5 \end{vmatrix} \]

\[ \cdots, \Delta l_{2N+1} = \begin{vmatrix} \mu_1 & \cdots & \mu_{N+1} \\ \vdots & \vdots & \vdots \\ \mu_{N+1} & \cdots & \mu_{2N+1} \end{vmatrix} , \quad (10) \]

\[ \Delta u_1 = \mu_0 - \mu_1, \quad \Delta u_3 = \begin{vmatrix} \mu_0 - \mu_1 & \mu_1 - \mu_2 \\ \mu_1 - \mu_2 & \mu_2 - \mu_3 \end{vmatrix} , \]

\[ \Delta u_5 = \begin{vmatrix} \mu_0 - \mu_1 & \mu_1 - \mu_2 & \mu_2 - \mu_3 \\ \mu_1 - \mu_2 & \mu_2 - \mu_3 & \mu_3 - \mu_4 \\ \mu_2 - \mu_3 & \mu_3 - \mu_4 & \mu_4 - \mu_5 \end{vmatrix} \]

\[ \cdots, \Delta u_{2N+1} = \begin{vmatrix} \mu_0 - \mu_1 & \cdots & \mu_N - \mu_{N+1} \\ \vdots & \vdots & \vdots \\ \mu_N - \mu_{N+1} & \cdots & \mu_{2N} - \mu_{2N+1} \end{vmatrix} . \quad (11) \]

The following results hold [8,9].

**Theorem 2.1.** The point \( \mu = (\mu_1, \ldots, \mu_M) \) is interior to \( DM \) if and only if all the Hankel determinants \( \Delta l_0, \Delta u_0, \Delta l_1, \Delta l_1, \ldots, \Delta l_M, \Delta l_M \) are positive.

**Theorem 2.2.** A sequence \( \mu_j, j = 0, \ldots, 2N \) of numbers is a system of moments of distribution on \([0, 1]\) iff the quadratic forms
$$f = \sum_{0}^{N} \mu_{i+j} x_i x_j, \quad F = \sum_{0}^{N-1} [\mu_{i+j+1} - \mu_{i+j+2}] x_i x_j,$$

are nonnegative.

**Theorem 2.3.** A sequence $\mu_j$, $j = 0, \ldots, 2N + 1$ of numbers is a system of moments of distribution on $[0, 1]$ iff the quadratic forms

$$g = \sum_{0}^{N} \mu_{i+j+1} x_i x_j, \quad G = \sum_{0}^{N} [\mu_{i+j} - \mu_{i+j+1}] x_i x_j,$$

are nonnegative.

**Remark 2.1.** It is well known that the moment sequence $\mu_j$, $j = 0, \ldots, 2N$ (or $j = 0, \ldots, 2N + 1$) allows a unique distribution $d\psi(x)$, whose first moments are equal to the assigned ones, with a finite number of increasing points $\xi_i$ iff one of the forms $f$ and $F$ (or $g$ and $G$) is singular. More precisely we have the following four cases depending on the first singular form (first Hankel’s determinant vanishing).

$$\Delta l_{2N} = 0 \quad (f \text{ singular}) \iff \xi_i, \quad i = 1, \ldots, N \in (0, 1), \ (N \text{ inner masses}),$$

$$\Delta u_{2N} = 0 \quad (F \text{ singular}) \iff \xi_i, \quad i = 1, \ldots, N - 1 \in (0, 1)(N - 1 \text{ inner masses}), \text{ and } \xi_0 = 0, \ \xi_N = 1,$$

$$\Delta l_{2N+1} = 0 \quad (g \text{ singular}) \iff \xi_i, \quad i = 1, \ldots, N \in (0, 1), \ (N \text{ inner masses}), \text{ and } \xi_0 = 0,$$

$$\Delta u_{2N+1} = 0 \quad (G \text{ singular}) \iff \xi_i, \quad i = 1, \ldots, N \in (0, 1), \ (N \text{ inner masses}), \text{ and } \xi_{N+1} = 1.$$

**Remark 2.2.** Next determinants (of order greater than $2N$ or $2N + 1$) are identically zero.

Once $(\mu_1, \ldots, \mu_M)$ are assigned then the extremes $M$th moment $\mu_M$ consistent with the first $M - 1$ components $(\mu_1, \ldots, \mu_{M-1})$ are indicated by $\mu_{M+}$ and $\mu_{M-}$ respectively. Such values are obtained as $\mu_M$ belongs to $\partial D^M$. Then the following relationships hold [8]
\[ \mu_M^+ - \mu_M^- \leq 2^{-2M+2}, \]  
(18)

\[ \mu_{2M} - \mu_{2M}^- = \frac{\Delta I_{2M}}{\Delta I_{2(M-1)}}. \]  
(19)

### 2.3. Geometrical meaning of Hankel determinants

To sequence of moments \( \mu_j, j = 0, 1, \ldots \) we associate the following positive sequence [10]

\[ \rho_{2n}(0) = \frac{\Delta I_{2n}}{\mu_2 \cdots \mu_{n+1} \atop \vdots \atop \mu_{n+1} \cdots \mu_{2n}} \quad n = 1, 2, \ldots \]  
(20)

As the infinite Hausdorff moment problem is determined then the following result holds [10]

\[ \rho(0) = \lim_{n \to \infty} \rho_{2n}(0) = 0. \]  
(21)

Once \( \mu_1, \ldots, \mu_M \) are assigned with \( M = 2n \), let \( \mu_{0,2n}^- \) the \( \mu_0 \) value so that \( \Delta I_{2n} = 0 \). From

\[ \Delta I_{2n} = \mu_0 \atop \vdots \atop \mu_{n+1} \cdots \mu_{2n} \quad + C, \]  
(22)

where \( C = C(\mu_1, \ldots, \mu_{2n}) \) and

\[ 0 = \mu_{0,2n}^- \atop \vdots \atop \mu_{n+1} \cdots \mu_{2n} \quad + C \]  
(23)

then

\[ \rho_{2n}(0) = \frac{\Delta I_{2n}}{\mu_2 \cdots \mu_{n+1} \atop \vdots \atop \mu_{n+1} \cdots \mu_{2n}} = \mu_0 - \mu_{0,2n}^- \quad n = 1, 2, \ldots \]  
(24)

represents the distance between \( \mu_0 \) and its lower bound \( \mu_{0,2n}^- \).
2.4. A general methodology

Our technique to obtain ME distribution existence conditions, entropy convergence, entropy decreasing, stability is as follows. Let us transform Eq. (5) in a system of differential equations by fixing \( l_1, \ldots, l_M \) and varying \( l_M \) in continuity within the range of its admissible values. Then \( k_0, \ldots, k_M \) coefficients are functions of \( l_M \). By differentiating both sides of Eq. (5) with respect to \( l_M \), one gets the system of differential equations

\[
\begin{align*}
\Delta l_{2M} & \cdot \begin{bmatrix}
\frac{d\lambda_0}{dl_M} \\
\vdots \\
\frac{d\lambda_M}{dl_M}
\end{bmatrix} = -
\begin{bmatrix}
0 \\
\vdots \\
0 \\
1
\end{bmatrix}
\] (25)
\end{align*}
\]

(where unknown moments \( \mu_{M+1}, \ldots, \mu_{2M} \) are obtained from Eq. (7)). Other variants of the method will be used, for instance fixing \( l_1, \ldots, l_M \) and varying \( l_0 \) in continuity within the range of its admissible values. Then we have

\[
\Delta l_{2M} \cdot \begin{bmatrix}
\frac{d\lambda_0}{dl_0} \\
\frac{d\lambda_1}{dl_0} \\
\vdots \\
\frac{d\lambda_M}{dl_0}
\end{bmatrix} = -
\begin{bmatrix}
1 \\
0 \\
\vdots \\
0
\end{bmatrix}
\] (26)

3. Existence of ME distribution

Considering Eq. (25) and Theorem 2.2 \( \lambda_M \) comes out as a decreasing monotonic function. By varying \( \mu_M \) within the bounded range of its admissible values, the determinant \( \Delta l_{2(M-1)} \) is bounded. As \( \mu_M \) tends to its lower or upper bound, then \( \lambda_M \to \pm \infty \). As a consequence also \( \Delta l_{2M} \to 0 \). We can argue that the lower \( \mu_M^- \) and upper \( \mu_M^+ \) bounds of \( \mu_M \), where \( \mu_M^- \leq \mu_M^+ \leq \mu_M^+ \), are determined by \( \Delta l_{2M} = 0 \).

Taking into account Remark 2.1 as \( \mu_M = \mu_M^- \) or \( \mu_M = \mu_M^+ \) the sequences \( \{\mu_1, \ldots, \mu_{M-1}, \mu_M^-\} \) or \( \{\mu_1, \ldots, \mu_{M-1}, \mu_M^+\} \) admit a unique distribution \( d\psi(x) \) and then \( \{\mu_1, \ldots, \mu_{M-1}, \mu_M^-\} \) and \( \{\mu_1, \ldots, \mu_{M-1}, \mu_M^+\} \) belong to the convex hull boundary. Thus \( \mu_M^- = \mu_M \) and \( \mu_M^+ = \mu_M \) hold. Then we have the following theorem.

**Theorem 3.1.** The necessary and sufficient conditions for the existence of a ME solution are identical to Hausdorff’s conditions for the finite moment problem, i.e.
the point \( \mu_1, \ldots, \mu_M \) be inner to moment space, or equivalently the Hankel determinants be positive.

This result has been previously obtained under more general constraints, by resorting to concept of I-divergence [2].

4. Entropy convergence

By definition \( H[f_0] \geq H[f_1] \geq \cdots \geq H[f_M] \geq \cdots \geq H[f] \). Taking into account Eq. (1), entropy-convergence means

\[
\lim_{M \to \infty} H[f_M] = H[f] \iff \lim_{M \to \infty} - \int_0^1 f_M(x) \ln f_M(x) \, dx
\]

\[
= - \int_0^1 f(x) \ln f(x) \, dx. \tag{27}
\]

Preliminarily, by combining Eqs. (1) and (4) we have

\[
H[f_M] = \sum_{j=0}^{M} \lambda_j \mu_j. \tag{28}
\]

We prove the following.

**Theorem 4.1.** Once the infinite sequence \( \{\mu_0, \mu_1, \ldots\} \) is assigned defining \( f(x) \) then ME approximant converges in entropy to \( f(x) \).

**Proof.** Let \( \mu_1, \ldots, \mu_M \) be assigned (without loss of generality \( M = 2n \) will be assumed) and let \( \mu_0 \) vary continuously. Then from Eq. (5) the Lagrange multipliers \( \lambda_j \) are functions of \( \mu_0 \). By differentiating Eq. (5) with respect to \( \mu_0 \) we have Eq. (26). From Eq. (28), by taking into account the first equation of Eq. (26), we have

\[
\frac{d}{d\mu_0} H[f_M] = \lambda_0 + \sum_{j=0}^{M} \mu_j \frac{d\lambda_j}{d\mu_0} = \lambda_0 - 1 \tag{29}
\]

and by Eqs. (24) and (26)

\[
\frac{d^2}{d\mu_0^2} H[f_M] = \frac{d\lambda_0}{d\mu_0} = - \frac{1}{\rho_{2M}(0)} < 0. \tag{30}
\]

Thus \( H[f_M] \) is a differentiable concave function of \( \mu_0 \).

Let us suppose that

\[
\lim_{M \to \infty} H[f_M] = H_{\lim} > H[f]. \tag{31}
\]
When, for given $\mu_0$, $M \to \infty$ then from a geometrical meaning of Hausdorff moment problem determinacy conditions we have $\lim_{M \to \infty} \mu_{0,2n} = \mu_0$ whereas, with $M$ fixed, by varying $\mu_0$, we have

$$\lim_{\mu_0 \to \mu_{0,2n}} H[f_M] = -\infty,$$

as from Eq. (14) $f_M(x)$ degenerates in a set of $n$ masses.

Let us consider the following sequences of ME solutions

1. $f_M(x)$ with moments $\mu_0, \ldots, \mu_M$
2. $f'_M(x)$ with moments $\mu_0 - \epsilon, \mu_1, \ldots, \mu_M$ with $\epsilon \in \mathbb{R}_+$ and $\mu_0 - \epsilon > \mu_{0,2n}$ so that $f'_M(x)$ exists and $H[f'_M] \leq H[f]$ holds. As $M \to \infty$ then from Eq. (21) $\epsilon \to 0$ so that the moment $\mu_0$ of $f_M(x)$ and $f'_M(x)$ differ in an arbitrarily small quantity, while their entropies differ in a finite quantity greater than $H_{\lim} - H[f]$.

Taking into account Eq. (31) a contradiction on the continuity of $H[f_M]$ as a function of $\mu_0$ is reached. Then $H_{\lim} = H[f]$ holds; therefore we have

$$\lim_{M \to \infty} H[f_M] = H[f].$$

Entropy-convergence has been previously proved for a bounded function $f(x)$ [3] and in the generalized Hausdorff moment problem [4]. The last authors proved how some properties of Shannon entropy, i.e. strict convexity, essential smoothness and coercivity lead to various distinct types of convergence.

We recall entropy-convergence entails $L_1$-norm convergence by resorting to concept of direct divergence. The result is as follows [11]

$$H[f_M] - H[f] = \int_0^1 f(x) \ln \frac{f(x)}{f_M(x)} \, dx \geq \frac{1}{4} \left( \int_0^1 |f(x) - f_M(x)| \, dx \right)^2$$

5. On entropy decreasing

We consider the entropy decreasing as one more moment is added. Let us fix $\mu_1, \ldots, \mu_{M-1}$ whilst $\mu_M$ varies in continuity. Then from Eq. (28) $\lambda_0, \ldots, \lambda_M$ and $H[f_M]$ are functions of $\mu_M$. By differentiating Eq. (28) with respect to $\mu_M$ and taking into account Eq. (25) we have

$$\frac{dH[f_M]}{d\mu_M} = \sum_{j=0}^{M} \mu_j \frac{d\lambda_j}{d\mu_M} + \lambda_M = \lambda_M$$

and by successive differentiation
Thus $H[f_M]$ is a differentiable concave function of $\mu_M$, assuming maximum value $H[f_{M-1}]$ as $\lambda_M = 0$. Thus we can write

$$H[f_0] \geq H[f_1] \geq \cdots \geq H[f_M] \geq \cdots \geq H[f].$$

Thus, adding one more moment doesn’t necessarily imply entropy decrease, unlike the case of Hamburger and Stieltjes moment problems where strict inequalities are allowed [12].

6. Stability analysis

Stability concerns the relation between $f_M(x)$ and $f'_M(x)$ characterized by $\mu = (\mu_0, \ldots, \mu_M)$ and $\mu' = (\mu_0 + \epsilon_0, \ldots, \mu_M + \epsilon_M)$, respectively, where $|\epsilon_j| \ll 1$ and $\mu' \in D^M$. In general, solving the infinite Hausdorff moment problem amounts to inverting a linear operator $A$ which associates to any function $f(x)$ the sequence of its all moments. From the relationship previous obtained $\lim_{M \to \infty} (\mu^+_M - \mu^-_M) = 0$ and from Eq. (25) $\lambda_M$ varies from $-\infty$ to $+\infty$ within the range $[\mu^-_M, \mu^+_M]$, it is evident the inverse operator $A$ is not continuous. In other words, if a solution $f_M(x)$ exists, then replacing $\mu = (\mu_0, \ldots, \mu_M)$ by $\mu' = (\mu_0 + \epsilon_0, \ldots, \mu_M + \epsilon_M)$ may totally destroy the solution of Eq. (5) even if max $|\epsilon_j|$ is arbitrarily small. Thus the Hausdorff moment problem is an ill-posed problem in the Hadamard sense.

As stressed in [13] the ultimate reason for the ill-posedness of Hausdorff moment problem is the lack of orthogonality of the sequence $1, x, x^2, \ldots, x^M, \ldots$. Heuristically speaking, different powers of $x$ differ very little from each other if $x$ is restricted in the interval $[0, 1]$ and the exponents are larger. Now we provide some quantitative relationships.

6.1. Multiplier $\lambda_M$ stability

We prove that, as the number $M$ of moments increases, the $\lambda_M$ multiplier computation determining $f_M(x)$, becomes unstable. Let $\mu^-_M$ and $\mu^+_M$ be the minimum and maximum values respectively of $\mu_M$, when the first $M - 1$ moments are fixed while $\mu_M$ varies within the range $[\mu^-_M, \mu^+_M]$. Fixed the first $M - 1$ moments, varying $\mu_M$ in continuity and taking into account Eqs. (25), (18) and (19) we have

$$-\frac{d^2 H[f_M]}{d\mu_M} = \frac{d\lambda_M}{d\mu_M} = -\frac{\Delta l_{2(M-1)}}{\Delta l_{2M}}.$$ (36)

Thus, adding one more moment doesn’t necessarily imply entropy decrease, unlike the case of Hamburger and Stieltjes moment problems where strict inequalities are allowed [12].

$$\frac{d^2 H[f_M]}{d\mu_M} = \frac{d\lambda_M}{d\mu_M} = -\frac{\Delta l_{2(M-1)}}{\Delta l_{2M}}.$$ (36)
from which
\[
\left| \frac{d \lambda_M}{d \mu_M} \right| > 2^{4M-2}
\]  
(39)

6.2. Behaviour of \( f_M(x) \) as the last moment \( \mu_M \) is varied

Let \( f_M(x; \lambda_j) \) the ME solution corresponding to the set of moments \( \mu_0, \ldots, \mu_M \) and \( f_M(x; \lambda_j + \Delta \lambda_j) \) the ME solution corresponding to the set of moments \( \mu_0, \ldots, \mu_{M-1}, \mu_M + \Delta \mu_M \). Let us consider the following quantity
\[
\frac{f_M(x; \lambda_j + \Delta \lambda_j) - f_M(x; \lambda_j)}{f_M(x; \lambda_j)} := \epsilon[f_M(x; \Delta \mu_M)],
\]  
(40)
representing the relative error on \( f_M(x) \) generated by the variation \( \Delta \lambda_j \) on \( \lambda_j \) given by a variation of \( \Delta \mu_M \) on \( \mu_M \). Because of the particular analytical form of \( f_M(x) \) we get
\[
\epsilon[f_M(x; \Delta \mu_M)] = \exp \left( - \sum_{j=0}^{M} \Delta \lambda_j \lambda^j \right) - 1 = \exp \left( - \Delta \mu_M \sum_{j=0}^{M} \frac{\Delta \lambda_j}{\Delta \mu_M} \lambda^j \right) - 1.
\]  
(41)

From Eq. (25) we obtain
\[
\frac{d \lambda_j}{d \mu_M} = - \frac{\Delta I_{2M}^{(j)} \lambda_j}{\Delta I_{2M}} , \quad j = 0, 1, \ldots, M,
\]  
(42)
where \( [\Delta I_{2M}^{(j)}] \) is obtained from matrix \( [\Delta I_{2M}] \) by substituting the \( j + 1 \)th column with the vector \( e_M = [0, \ldots, 0, 1]^T \). Then we have
\[
\epsilon[f_M(x; \Delta \mu_M)] = \exp \left( \frac{\Delta \mu_M}{\Delta I_{2M}} \sum_{j=0}^{M} \Delta I_{2M}^{(j)} \lambda^j \right) - 1 = \exp \left( \frac{\Delta \mu_M}{\Delta I_{2M}} P_M(x) \right) - 1;
\]  
(43)
where
\[
P_M(x) = \begin{bmatrix}
\mu_0 & \cdots & \mu_{M-1} & 1 \\
\mu_1 & \cdots & \mu_M & x \\
\vdots & \vdots & \vdots & \vdots \\
\mu_M & \cdots & \mu_{2M-1} & x^M
\end{bmatrix}
\]  
(44)
is, up to a constant factor, the orthogonal polynomial of degree \( M \) defined by the first \( 2M - 1 \) moments of \( f_M(x) \).

From Eq. (44), by well known properties of the zeros of orthogonal polynomials, there exist \( M \) points where \( \epsilon[f_M(x; \Delta \mu_M)] = 0 \) holds. ME solutions \( f_{M-1}(x) \) and \( f_M(x) \), having the same first \( M - 1 \) moments, have exactly \( M \)
common points. This property can be considered a characteristic of the ME solutions as, in general, two continuous functions having the same first \(M - 1\) moments have at least \(M\) common points [14].

From Eq. (43), by using Taylor expansion, we get

\[
\epsilon(f_M(x; \Delta \mu_M)) \simeq \frac{\Delta \mu_M}{\Delta l_{2M}} P_M(x) = \frac{\Delta \mu_M}{\mu_M} \frac{P_M(x)}{\Delta l_{2M}} \mu_M
\]  

(45)

Thus the relative error on \(\mu_M\) is amplified by the factor

\[
\mu_M \frac{|P_M(x)|}{\Delta l_{2M}}
\]

(46)

representing a condition number of \(f_M(x)\) as \(\mu_M\) is varied.

From Eq. (45) we observe that, as the number of moments increases, a perturbation on the last moment determines a null variation of the function in an increasing number \(M\) of points, located exactly in the zeros of the orthogonal polynomial \(P_M(x)\).

6.3. Relative error estimate

By taking into account the following relationship [8]

\[
\int_0^1 P_M^2(x) f_M(x) \, dx = \Delta l_{2M} \Delta l_{2(M-1)}
\]  

(47)

and Eqs. (47), (45) and (38) we have

\[
\int_0^1 |\epsilon(f_M(x; \Delta \mu_M))|^2 f_M(x) \, dx \simeq \left( \frac{\Delta \mu_M}{\Delta l_{2M}} \right)^2 \int_0^1 P_M^2(x) f_M(x) \, dx
\]

\[
= \left( \frac{\Delta \mu_M}{\mu_M} \right)^2 \Delta l_{2M} \Delta l_{2(M-1)}
\]

\[
= \frac{(\Delta \mu_M)^2}{\mu_M^2 - \mu_{2M}^2} > \frac{(\Delta \mu_M)^2}{\mu_M^2 - \mu_{2M}^2} \geq \frac{(\Delta \mu_M)^2}{2^{-4M+2}} = (\Delta \mu_M)^2 2^{4M-2}.
\]

(48)

Here \(f_M(x; \Delta \mu_M)\) represents the calculated \(f_M(x)\), so that \(f^{\text{calc}}_M(x) = f_M(x; \Delta \mu_M)\). Then from Eq. (48) we have the weighted-square-relative error estimate

\[
\int_0^1 \left| \frac{f^{\text{calc}}_M(x) - f_M(x)}{f_M(x)} \right|^2 f_M(x) \, dx > (\Delta \mu_M)^2 2^{4M-2},
\]  

(49)

where

\[
\Delta \mu_M = \left| \mu_M - \int_0^1 x^M f^{\text{calc}}_M(x) \, dx \right|
\]

(50)

is essentially related to accuracy of the algorithm used in minimizing Eq. (6).
6.4. Entropy decreasing estimate

Taking into account Eq. (33), for high \( M \) values we can assume \( H[f_M] \approx H[f_{M-1}] \). Let \( \mu_M^* \) the \( \mu_M \) value so that \( \lambda_M = 0 \) and then
\[
dH[f_M(\mu_M^*)]/d\mu_M = 0.
\]
Thus from Eq. (37) we have \( H[f_M(\mu_M^*)] = H[f_{M-1}] \).

From Taylor expansion about \( \mu_M^* \) and taking into account Eqs. (18), (19), (36) and (38) we have
\[
-H[f_M(\mu_M^* + \Delta\mu_M)] + H[f_M(\mu_M^*)] = -\frac{1}{2} (\Delta\mu_M)^2 \frac{d^2 H[f_M(\xi_M)\mu_M^*]}{d\mu_M^2}
\]
\[
= \frac{1}{2} (\Delta\mu_M)^2 \frac{\Delta l_{2(M-1)}}{\Delta l_{2M}}
\]
\[
= \frac{(\Delta\mu_M)^2}{2(\xi_{2M} - \mu_{2M})} > \frac{(\Delta\mu_M)^2}{2(\mu_{2M}^+ - \mu_{2M}^-)} = (\Delta\mu_M)^2 2^{4M-3}
\]
where \( \mu_{\Delta M} < \xi_M < \mu_M^+ \) and \( \mu_{2M}^- < \xi_{2M} < \mu_{2M}^+ \), while \( \Delta l_{2(M-1)} \) and \( \Delta l_{2M} \) are evaluated substituting \( \mu_M \) by \( \xi_M \).

6.5. Averages estimate

Rather than \( f(x) \) itself, one is most often interested in averages of some functions \( F(x) \) over the distribution to obtain
\[
\langle F \rangle = \int_0^1 F(x)f(x) \, dx
\]  
(52)

Then Eq. (52) is approximated by
\[
\langle F_M \rangle = \int_0^1 F(x)f_M(x; \lambda_j) \, dx
\]  
(53)

(the same notation as Section 6.2 is adopted). Here we assume \( F(x) \) a bounded function, so that \( |F(x)| \leq K \).

We thus desire an estimate of the difference between \( \langle F_M \rangle \) and the computed average \( \langle F_M^{\text{calc}} \rangle \) with \( f_M(x; \lambda_j) \) replaced by \( f_M(x; \lambda_j + \Delta\lambda_j) \). The absolute difference satisfies the following inequalities
\[
\Delta F_M = |\langle F_M \rangle - \langle F_M^{\text{calc}} \rangle|
\]
\[
= \left| \int_0^1 F(x)[f_M(x; \lambda_j) - f_M(x; \lambda_j + \Delta\lambda_j)] \, dx \right| \leq \int_0^1 |F(x)| \cdot |f_M(x; \lambda_j) - f_M(x; \lambda_j + \Delta\lambda_j)| \, dx
\]
\[ - \int f_M(x; \lambda_j + \Delta \lambda_j) \, dx \]
\[ \leq K \int_0^1 [f_M(x; \lambda_j) - f_M(x; \lambda_j + \Delta \lambda_j)] \, dx \]
\[ \simeq K \frac{\Delta \mu_M}{\Delta I_{2M}} \int_0^1 |P_M(x)| f_M(x; \lambda_j) \, dx \]
(54)

Taking into account Eq. (34) for increasing \( M \) values (\( f(x) \) replaced by \( f_M(x; \lambda_j + \Delta \lambda_j) \)), from Eq. (54) we have
\[ \Delta F_M \simeq o(\Delta \mu_M) \]  
(55)

From Eqs. (39), (49), (51) and (55) we conclude

1. for increasing \( M \) values the \( \lambda_j \) multipliers calculation becomes highly unstable;
2. for moderate \( M \) values the approximant \( f_M(x) \) and entropy \( H[f_M] \) can be reliably estimated.

7. Conclusions

Some analytical and computational aspects, as existence condition, entropy convergence, entropy decreasing, stability concerning Hausdorff moment problem in the framework of ME have been reviewed through a unified approach. Results known in literature and some new ones concerning the stability of the computed solution have been illustrated.

References