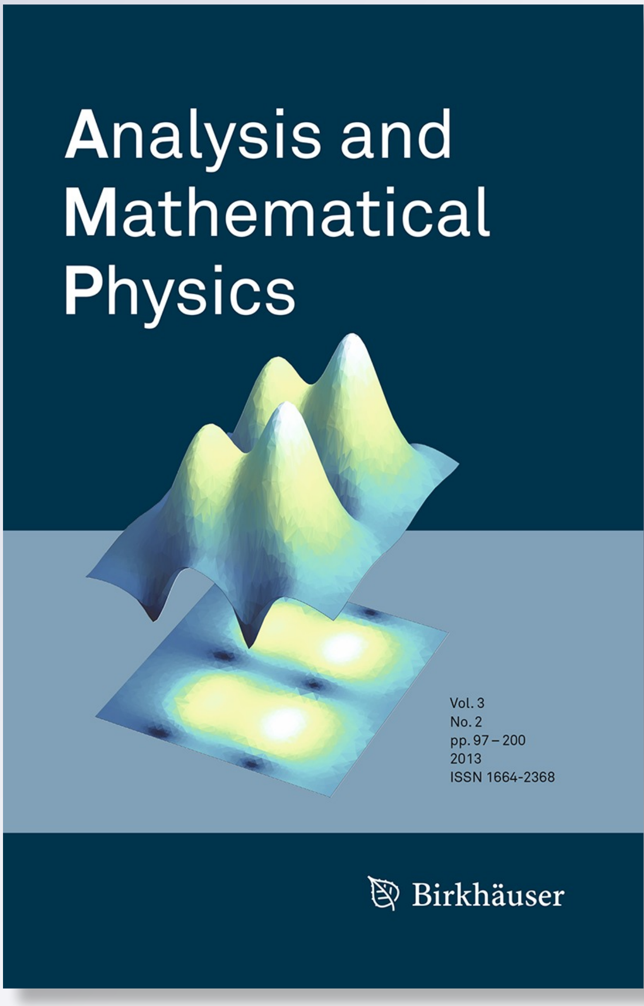
Multivariate truncated moments problems and maximum entropy

Calin -Grigore Ambrozie

Analysis and Mathematical Physics

ISSN 1664-2368
Volume 3
Number 2

Anal.Math.Phys. (2013) 3:145-161
DOI 10.1007/s13324-012-0052-3



Springer

Anal.Math.Phys. (2013) 3:145–161
DOI 10.1007/s13324-012-0052-3

Multivariate truncated moments problems and maximum entropy

Calin-Grigore Ambrozie

Received: 14 May 2012 / Revised: 25 December 2012 / Accepted: 27 December 2012 /
Published online: 9 January 2013

Abstract We characterize the existence of the Lebesgue integrable solutions of the truncated problem of moments in several variables on unbounded supports by the existence of some maximum entropy—type representing densities and discuss a few topics on their approximation in a particular case, of two variables and 4th order moments.

Keywords Moments problem · Representing measure · Entropy

Mathematics Subject Classification (2010) Primary 44A60; Secondary 35A15

1 Introduction

In this work we consider the problem of moments in the following context. Let $T \subset \mathbb{R}^n$ be a closed subset, where $n \in \mathbb{N}$ is fixed. Let $I \subset (\mathbb{Z}_+)^n$ be finite such that $0 \in I$, where $\mathbb{Z}_+ = \mathbb{N} \cup \{0\}$. Fix a set $g = (g_i)_{i \in I}$ of real numbers g_i with $g_0 = 1$. The problem under consideration is to establish if there exist (classes of) Lebesgue measurable

Supported by the Czech grant IAA100190903 GAAV (RVO: 67985840) and the Romanian grant CNCS UEFISCDI, no. PN-II-ID-PCE-2011-3-0119.

C.-G. Ambrozie (✉)
Institute of Mathematics of the Czech Academy, Zitna 25, 115 67 Prague 1, Czech Republic
e-mail: ambrozie@math.cas.cz

C.-G. Ambrozie
Institute of Mathematics "Simion Stoilow"-Romanian Academy, P.O. Box 1-764,
014700 Bucharest, Romania

functions $f \geq 0$ a.e. (almost everywhere) on T, such that $\int_T |t^i| f(t)\,dt < \infty$ and

$$\int_T t^i f(t)\,dt = g_i \ \ (i \in I) \tag{1}$$

and find such solutions f. As usual $dt = dt_1 \dots dt_n$ and $t^i = t_1^{i_1} \cdots t_n^{i_n}$ for any multiindex $i = (i_1, \dots, i_n) \in \mathbb{Z}_+^n$ where $t = (t_1, \dots, t_n)$. In this case we call f a *representing density* for g, and g_i the *moments* of f. In general T is unbounded and usually $I = \{i : |i| \leq 2k\}$ for $k \in \mathbb{N}$, where $|i| = i_1 + \cdots + i_n$.

Generally a problem of moments [5,37], called also T*-problem of moments* when T is given (Curto and Fialkow [15]), is concerned with the existence of an arbitrary Borel measure $\mu \geq 0$ supported on T such that $\int_T t^i d\mu(t) = g_i$ for $i \in I$, in which case one calls μ a *representing measure* of g. The feasibility of (1) characterizes the dense interior of the convex cone of all data g having representing measures, provided that all $t \in T$ are density points and I is a union of intervals $[0, i] := \{j \in \mathbb{Z}_+^n : 0 \leq j_k \leq i_k,\ 1 \leq k \leq n\}$, see (Theorems 5, 6, [3], and Junk, Theorem A.1, [23] in a slightly different context). For our purpose here we only require that the Lebesgue measure of T be $\neq 0$.

Author's main contributions are contained in the statements 4–9. In particular, by Corollary 6 for example, for each fixed $\epsilon > 0$ we characterize the feasibility of (1) by the existence of a (unique) f_* minimizing $\int_T f \ln f\,dt + \epsilon \int_T \|t\|^{2k+2} f(t)\,dt$ amongst all solutions, which is equivalent to the existence of a (unique) vector $\lambda^* = (\lambda_i^*)_{|i| \leq 2k}$ maximizing the associated Lagrangian $L(\lambda) = L_\epsilon(\lambda) = \sum_{|i| \leq 2k} g_i \lambda_i - \int_T e^{\sum_{|i| \leq 2k} \lambda_i t^i - \epsilon \|t\|^{2k+2}} dt$, in which case $f_*(t) = e^{\sum_{|i| \leq 2k} \lambda_i^* t^i - \epsilon \|t\|^{2k+2}}$ where $\|t\| = (\sum_{j=1}^n t_j^2)^{1/2}$. The more general formulation of the main result Theorem 4 aims to cover also other cases like T =compact with $\epsilon = 0$ [31].

Maximizing the Boltzmann–Shannon's *entropy* $H(f) = -\int_{\mathcal{T}} f \ln f\,d\mu$ on a probability space $(\mathcal{T}, \mu)$ subject to various restrictions $\int_{\mathcal{T}} a_i f\,d\mu = g_i \ (i \in I)$ is a well-known principle in statistical mechanics and information theory [13,20,23,32]. The maximum of H is attained on the unbiased probability distribution f_* on a partial knowledge, of the prescribed average values g_i of some random variables [9,13,20]. Typically f_* is obtained by maximizing a function L (the Lagrangian) convex conjugate to $-H$ [10,11,27,33,36], which leads to characterizations (sup/max$L < \infty$) of the feasibility of the primal problem—in our case (1). One may consider more general measures $\mu \geq 0$ or functionals like $H(f) = -\mathrm{tr}(f \ln f)$, $\mathrm{tr}(\ln f)$ where $f =$ positive definite matrix for noncommutative moments (Theorems 2,3, [4]), [7].

While the case T =compact was known long before (Lewis [31]), the similar problems with unbounded support T (or unbounded moments a_i) are usually difficult and still studied, see Abramov [1], Borwein [9], Junk [23], Hauck et al. [27] and others [18,22,24,30,39]. We mention that the feasibility of (1) has been characterized in Blekherman, Lasserre [12], avoiding entropy maximization but also in Lagrangian terms. If T is unbounded, Corollary 6 cannot be improved to $\epsilon = 0$: there are examples of realizable, but degenerate data g such that the constrained H-maximization fails for $(\mathcal{T}, \mu) = (\mathbb{R}^n, dt)$, see Junk and co-workers [22,27]. For $H(f) = -\int_T f \ln f\,dt$, the maximization of $L(\lambda)$ $(= L_0(\lambda) \in [-\infty, \infty))$ always holds, at a unique point λ^*—using for instance (Corollary 2.6, [10]), see also [23,27]. It follows, by means of Fatou's lemma, for $I = \{i : |i| \leq 2k\}$, that $|t^i|e^{\sum_{|j|\leq 2k} \lambda_j^* t^j} \in L^1(T, dt)$ for all $|i| \leq 2k$, $\int_T t^i e^{\sum_{|j|\leq 2k} \lambda_j^* t^j} dt = g_i$ $(|i| < 2k)$ but the equality may fail for $|i| = 2k$. Namely the dual attainment $\sup L = \max L$ does hold, but primal attainment $\sup_{f\in(1)} H(f) = \max_{f\in(1)} H(f)$ is also a difficult topic if $a_i(t)$ (for instance t^i) are not in the dual of $L^1(T)$. For these matters we refer the reader to [22,23,27,32] and for applications to Boltzmann equations we mention also [6,14,18,24,39].

Originated in works by Stieltjes, Hausdorff, Hamburger and Riesz, the area of moments problems saw extensive development in many directions, that we do not attempt to cover. There exist also other approaches to the multivariate moments problems, by operator theoretic or convexity methods [16,17,19,35,38,40], in particular a truncated version of Riesz–Haviland's theorem [15], see also [25,34] for other results, related to sums-of-squares representations of positive polynomials or polynomial optimization theory. These interesting topics are beyond the goal of the present paper, that is focused on the H/L -maximization.

2 Main results

Fix T, I and g as stated in the Introduction. For any measurable space $\mathcal{T}$ endowed with a σ-finite measure $\mu \geq 0$ and $1 \leq p \leq \infty$, the notations $L^p(\mathcal{T}, \mu)$, $L^p_+(\mathcal{T}, \mu)$ (sometimes, $L^p(\mu)$, $L^p_+(\mu)$) have the usual meaning. We repeat below an argument from (Theorem 2.9, [10]), adapted to our case.

Lemma 1 (See [10]) *Let $\mu \geq 0$ be a finite measure on $\mathcal{T}$. Let $x \in L^1_+(\mu)\backslash\{0\}$, and $a_i \in L^1(\mu)$ $(i \in I)$ be a finite set of functions such that $\int_{\mathcal{T}} |a_i| x d\mu < \infty$ for all i and $(a_i)_i$ are linearly independent on any subset of positive measure. Then there is a sequence $(y_k)_{k\geq k_0} \subset L^\infty(\mu)$ such that $x_k := \min(x, k) + y_k \geq 0$ a.e., $\int_{\mathcal{T}} a_i x_k d\mu = \int_{\mathcal{T}} a_i x\, d\mu$ for all $i \in I$, $|y_k| \leq x$ and $y_k \to 0$ a.e.*

Proof Set $z_k = \min(x, k)$ for $k \geq 1$. Using $\{x > 0\} = \cup_{l\geq 1}\{x \geq 1/l\}$, we find a $\delta \in (0, 1)$ and $\mathcal{T}_* \subset \mathcal{T}$ with $\mu(\mathcal{T}_*) > 0$ such that $x(t) \geq \delta$ a.e. on $\mathcal{T}_*$. The linear map $A : L^\infty(\mathcal{T}_*) \to \mathbb{R}^N$ $(N = \text{card}\, I)$, $Ay = (\int_{\mathcal{T}_*} a_i y\, d\mu)_i$ is surjective for otherwise there is a $(\lambda_i)_i \neq 0$ orthogonal to its range, such that $\sum_i \lambda_i \int_{\mathcal{T}_*} a_i y\, d\mu = 0 \forall y$, whence $\sum_i \lambda_i a_i = 0$ a.e. on $\mathcal{T}_*$ that is impossible. Since A has closed range, there is a c such that $\inf_{w\in\ker A} \|y - w\|_\infty \leq c\|Ay\| \forall y \in L^\infty(\mathcal{T}_*)$. By Lebesgue's theorem of dominated convergence, $\lim_k \int_{\mathcal{T}} a_i z_k d\mu = \int_{\mathcal{T}} a_i x\, d\mu$ for all i. There are $y_k \in L^\infty(\mathcal{T})$ with $\text{supp}\, y_k \subset \mathcal{T}_*$ such that $\int_{\mathcal{T}} a_i y_k d\mu = \int_{\mathcal{T}} a_i (x - z_k)\, d\mu$ and, since $Ay_k \to 0$, we can choose them such that $\|y_k\|_\infty \to 0$. For large k, $\|y_k\|_\infty \leq \delta/2$. On $\mathcal{T}_*$, $x \geq \min(x, k) \geq \delta > \delta/2 \geq |y_k|$. Then $x_k \geq 0$ a.e. □

Fenchel duality deals with minimizing convex functions $\varphi : X \to (-\infty, \infty]$ over convex subsets of locally convex spaces X, in connection with the dual problem of maximizing $-\varphi^*$ where φ^* is the *convex conjugate* of φ, called also its *Legendre-Fenchel transform* [10,11,27,33,36]; φ must be *proper* $(\varphi \not\equiv \infty)$. Letting the *effective domain* of φ be $\text{dom}\, \varphi = \{x \in X : \varphi(x) < \infty\}$, φ^* is defined on the dual of X by

$\varphi^*(x^*) = \sup\{\langle x, x^*\rangle - \varphi(x) \,:\, x \in \operatorname{dom}\varphi\}$. Typically, $\inf\varphi = \sup(-\varphi^*)$. Briefly speaking, we set $\varphi(x) = -H(x)$ if $x \geq 0$ satisfies the equations of moments, and $\varphi(x) = +\infty$ outside the set of solutions. Then φ is convex conjugate to $\varphi^*(x^*) = \ln\int_T e^{\sum_i \lambda_i a_i}\,d\mu - \sum_i g_i\lambda_i$ for $x^* = \sum_i \lambda_i a_i$, and $\varphi^*(x^*) = +\infty$ otherwise. Thus $\operatorname{dom}\varphi^*$ is the linear span of the a_i's and (if $a_0 \equiv 1$) the Lagrangian $l := -\varphi^*|_{\operatorname{dom}\varphi^*}$ is given by $\lambda \mapsto -\ln\int_T e^{\sum_{i\in I\setminus\{0\}}\lambda_i(a_i-g_i)}d\mu$. Maximizing l or L are equivalent problems. We rely on Borwein and Lewis' results [10] concerned with L, providing dual attainment in a point λ^*. The equality $\inf\varphi = \sup(-\varphi^*)$ becomes here $P = D$. Although under different hypotheses, L is analogous to the dual function ψ from Hauck, Levermore and Tits (Section 4.1, [27]), and would fit the case when $\operatorname{dom} L \cap \partial(\operatorname{dom} L) = \emptyset$ in Junk [23] except we do not have here a distinguished moment a_m such that $\lim_{\|t\|\to\infty} \frac{|a_i(t)|}{1+a_m(t)} = 0\ (i \neq m)$.

The following Borwein and Lewis' result from [10] is the main Fenchel theoretic tool to be used later on in the proof of our Theorem 4.

Theorem 2 (Corollary 2.6, [10]) *Let $\mathcal{T}$ be a space with finite measure $\mu \geq 0$, $1 \leq p \leq \infty$ and $a_i \in L^q(\mu)$, $g_i \in \mathbb{R}$ for $i \in I$ (=finite) where $\frac{1}{p} + \frac{1}{q} = 1$. Let $\phi : \mathbb{R} \to (-\infty, \infty]$ be proper, convex and lower semicontinuous, with* $(0,\infty) \subset$ *$\operatorname{dom}\phi$. Suppose there exist $x \in L^p(\mu)$ with $x(t) > 0$ a.e. such that $\phi\circ x \in L^1(\mu)$ and $\int_{\mathcal{T}} a_i(t)\,x(t)\,d\mu(t) = g_i$ for $i \in I$. Then the values $P \in [-\infty, \infty)$ and $D \in [-\infty, \infty]$ defined respectively by*

$$P = \inf\left\{\int_{\mathcal{T}} \phi(x(t))\,d\mu(t) : x \in L^p(\mu),\ x \geq 0 \text{ a.e.},\ \phi\circ x \in L^1(\mu), \int_{\mathcal{T}} a_i x\,d\mu = g_i\ \forall i\right\}$$

and

$$D = \max\left\{\sum_{i\in I} g_i\lambda_i - \int_{\mathcal{T}} \phi^*\left(\sum_{i\in I}\lambda_i a_i(t)\right) d\mu(t) : \lambda_i \in \mathbb{R},\ \phi^*\circ\sum_{i\in I}\lambda_i a_i \in L^1(\mu)\right\}$$

are equal, $-\infty \leq P = D < \infty$ and the maximum D is attained.

Remark 3 (a) Let ϕ be defined by $\phi(x) = x\ln x$ for $x > 0$, $\phi(0) = 0$ and $\phi(x) = +\infty$ for $x < 0$. Then ϕ is proper, convex, lower semicontinuous, bounded from below, with effective domain $[0,\infty)$ and its convex conjugate is $\phi^*(y) = e^{y-1}$ for all $y \in \mathbb{R}$; use to this aim that $\phi^*(y) = \sup_{x\geq 0}(xy - x\ln x)$.

(b) For the integrand ϕ defined at (a) and $(\lambda_i)_{i\in I} = 0$, the constant function $(\phi^*\circ\sum_{i\in I}\lambda_i a_i)(t) \equiv \phi^*(0)$ is in $L^1(\mu)$. Thus for any data a_i, g_i verifying the hypotheses of Theorem 2, we obtain that $-\infty < P = D < \infty$.

(c) Let $x \in L^1_+(\mu)$ with $x\ln x \in L^1(\mu)$ and $y_k \in L^1(\mu)$ $(k \geq 1)$ such that $x_k := \min(x,k) + y_k \geq 0$ a.e., $|y_k| \leq x$ and $y_k \to 0$ a.e. as $k \to \infty$. By Lebesgue's dominated convergence theorem, $\lim_k \int x_k \ln x_k d\mu = \int x\ln x\,d\mu$, since on $\{t : x_k(t) \geq 1\}$, $x_k \leq 2x \Rightarrow |x_k\ln x_k| \leq |2x\ln(2x)|$ while on $\{t : x_k(t) < 1\}$, $|x_k\ln x_k| \leq 1/e$; hence $|x_k\ln x_k| \leq |2x\ln x + (2\ln 2)x| + 1/e \in L^1(\mu)$.

In Theorem 4 the choice of the norm on $\mathbb{R}^n$ is unimportant. We call a function a on T independent of $(t^i)_{i\in I\setminus\{0\}}$ if there are no subsets $Z\subset T$ of positive measure and constants $(c_i)_{i\in I\setminus\{0\}}$ such that $a=\sum_{i\in I\setminus\{0\}} c_i t^i$ on Z.

Theorem 4 *Let $T\subset\mathbb{R}^n$ be closed, $I\subset\mathbb{Z}_+^n$ finite, $0\in I$ and $g=(g_i)_{i\in I}$ a set of numbers with $g_0=1$. Set $m=\max_{i\in I}|i|$. Let a, ρ be measurable functions on T, $0<a,\ \rho<\infty$ a.e. such that $\int_T e^{\frac{\|t\|^m+1}{\alpha\, a(t)}}\rho(t)\,dt<\infty$ for all $\alpha>0$, and a is independent of $(t^i)_{i\in I\setminus\{0\}}$. The statements (a), (b), (c) are equivalent:*

(a) There exist functions $f\in L^1_+(T,dt)$ such that $\int_T |t^i|f(t)dt<\infty$ and

$$\int_T t^i f(t)\,dt = g_i\ \ (i\in I); \tag{2}$$

(b) There exists a particular solution $f_*\in L^1_+(T,dt)$ of problem 2, maximizing the entropy functional $H=H_{\rho,a}: L^1_+(T,dt)\to[-\infty,\infty)$ given by

$$H(f) = -\int_T \left(\frac{af}{\rho}\ln\frac{af}{\rho}\right)\rho\,dt$$

amongst all solutions;

(c) The Lagrangian function $L=L_{\rho,a,g}:\mathbb{R}^N\to[-\infty,\infty)$ ($N=\mathrm{card}\, I$) associated to the functional H and the Eq. 2, given by

$$L(\lambda) = \sum_{i\in I} g_i\lambda_i - \int_T e^{\sum_{i\in I}\lambda_i t^i/a(t)-1}\rho(t)\,dt \quad (\lambda=(\lambda_i)_{i\in I}),$$

is bounded from above and attains its supremum in a point $\lambda^*=(\lambda_i^*)_{i\in I}$.

In this case f_* and λ^* are uniquely determined, $-H(f_*)=L(\lambda^*)$ and

$$f_*(t) = a(t)^{-1}\rho(t)\,e^{\sum_{i\in I}\lambda_i^* t^i/a(t)-1}\quad (t\in T),$$

in particular $H\not\equiv-\infty$ on the set of all solutions of 2, and

$$\int_T t^i\, e^{\frac{1}{a(t)}\sum_{j\in I}\lambda_j^* t^j-1}a(t)^{-1}\rho(t)\,dt = g_i\quad (i\in I).$$

Proof Let $a_i(t)=t^i/a(t)$ for $i\in I$ and $t\in T$. The condition on ρ and a shows that the measure $\mu:=\rho\,dt$ on T is finite and, by means of the inequalities: $|t_j|\le\|t\|$ $(:=(\sum_{j=1}^n t_j^2)^{1/2})$ for $1\le j\le n$,

$$|t^i| = |t_1^{i_1}\cdots t_n^{i_n}| \le \|t\|^{i_1+\cdots+i_n} \le (\|t\|^m+1)^{|i|/m} \le \|t\|^m+1 \tag{3}$$

and $\sum_{i\in I}\lambda_i a_i(t)\le\sum_{i\in I}|\lambda_i|\cdot\frac{\|t\|^m+1}{a(t)}$, that for every $\lambda=(\lambda_i)_{i\in I}\in\mathbb{R}^N$

$$g(\lambda):=\int_T e^{\sum_{i\in I}\lambda_i a_i(t)-1}d\mu(t)<\infty. \tag{4}$$

By writing $se^{s/\alpha}\le e^{\beta s/\alpha}$ for large β ($\ge\alpha/e+1$) and $s:=(\|t\|^m+1)/a(t)$,

$$\int_T \frac{\|t\|^m+1}{a(t)}e^{\frac{\|t\|^m+1}{\alpha a(t)}}\,d\mu(t)<\infty\quad(\alpha>0). \tag{5}$$

Then for every $\lambda=(\lambda_i)_{i\in I}$, by the inequalities (3) again,

$$\int_T(\|t\|^m+1)a(t)^{-1}e^{\sum_{i\in I}\lambda_i a_i(t)-1}d\mu(t)<\infty. \tag{6}$$

Hence $\int_T a_i(t)e^{\sum_{i\in I}\lambda_i a_i(t)-1}d\mu(t)<\infty$, in particular $a_i\in L^1(T,\mu)$ for $i\in I$. Any of the statements (a)–(c) implies that the Lebesgue measure of T is strictly positive (finite or not), due to the condition $g_0=1$. Then for every $f\in L^1_+(T,dt)$, by Jensen's inequality for the function $\phi(x):=x\ln x$ ($x\ge 0$),

$$H(f)=-\mu(T)\int_T\phi\left(\frac{af}{\rho}\right)\frac{d\mu}{\mu(T)}\le-\mu(T)\,\phi\left(\int_T\frac{af}{\rho}\frac{d\mu}{\mu(T)}\right)\le\mu(T)/e<\infty.$$

(a) $\Rightarrow$ (c). Suppose that problem (2) has a solution f. The function $x:=af/\rho$ then satisfies $\int_T|a_i|x\,d\mu<\infty$ and $\int_T a_i x\,d\mu=g_i$ for $i\in I$. By the original version (Theorem 2.9, [10]) of Lemma 1 (if $x_k=\max(x,k)+\frac{1}{k}+y_k$), there are functions $\tilde{x}\in L^\infty(T,\mu)$, $\tilde{x}>0$ μ-a.e. on T, such that $\int_T a_i(t)\tilde{x}(t)d\mu=g_i$ ($i\in I$). Here $L^\infty(T,\mu)=L^\infty(T,dt)$ since μ is equivalent to dt on T. For such $\tilde{x}$, the function $\phi\circ\tilde{x}=\tilde{x}\ln\tilde{x}$ belongs to $L^\infty(T)$, and hence, to $L^1(T,\mu)$. Then we can use Theorem 2 for $\phi(x)=x\ln x$ and $p=\infty$, see Remark 3, (a). Let $P=\inf_x\int_T x\ln x\,d\mu$ over the set of all $x\in L^\infty_+(T)$ such that

$$\int_T a_i x\,d\mu=g_i,\ i\in I \tag{7}$$

and $D:=\sup L$. Then $-\infty<P=D<\infty$ with attainment in the dual problem, see Remark 3, (b). Therefore, $L(\lambda)=\sum_{i\in I}g_i\lambda_i-\int_T e^{\sum_{i\in I}\lambda_i a_i(t)-1}d\mu(t)$ is bounded from above on $\mathbb{R}^N$ and its supremum D is attained. (c) $\Rightarrow$ (b). Assume there is a λ^* such that $L(\lambda^*)=\max L$. As expected, we will derivate under the integral to show that $x_*(t):=e^{\sum_{i\in I}\lambda_i^* a_i(t)-1}$ satisfies (7) and moreover maximizes $H_\mu(x):=-\int_T x\ln x\,d\mu$ amongst all solutions from $L^1_+(T,\mu)$. Firstly, by (4),

$\int_T x_*(t)d\mu(t) = g(\lambda^*)$. By (3) and (6), $\int_T |a_i|x_* d\mu < \infty$ $(i \in I)$. For any λ we have $L(\lambda) \le L(\lambda^*)$, that is, by (4),

$$g(\lambda^*) \le g(\lambda) + \sum_{i\in I} g_i(\lambda_i^* - \lambda_i). \tag{8}$$

Fix $j \in I$, let $\varphi(t) = \pm a_j(t)$ and set $v = (v_i)_{i\in I}$ where $v_i = \pm\delta_{i\,j}$ = Kronecker's symbol (the signs agree). For any $\varepsilon > 0$, set $\lambda_\varepsilon = \lambda^* + \varepsilon v$, namely $\lambda_\varepsilon = (\lambda_{\varepsilon i})_{i\in I}$ where $\lambda_{\varepsilon j} = \lambda_j^* \pm \varepsilon$ and $\lambda_{\varepsilon i} = \lambda_i^*$ for $i \ne j$. Let $F_\varepsilon(t) = \frac{1}{\varepsilon}x_*(t)(1 - e^{\varepsilon\varphi(t)})$. Note that

$$\lim_{\varepsilon\to 0} F_\varepsilon(t) = -\varphi(t)x_*(t) \tag{9}$$

and $x_* e^{\varepsilon\varphi} = e^{\sum_{i\in I}\lambda_i^* a_i - 1 + \varepsilon(\pm a_j)} = e^{\sum_{i\in I}\lambda_{\varepsilon i} a_i - 1}$. Then by (4) and (8),

$$\int_T F_\varepsilon(t)d\mu(t) = \frac{g(\lambda^*) - g(\lambda_\varepsilon)}{\varepsilon} \le \mp g_j. \tag{10}$$

By the estimates (3), we may let $y = \varphi(t)$ and $z = (\|t\|^m + 1)/a(t)$ in the inequality: $e^{-z}\frac{1-e^{\varepsilon y}}{\varepsilon} \ge -|y|$ where $z > 0$, y is real, $|y| \le z$ and $\varepsilon < 1$. Hence $F_\varepsilon(t) \ge -x_*(t)|\varphi(t)| \cdot e^{(\|t\|^m+1)/a(t)}$. The right hand side is in $L^1(T, \mu)$ by the estimates: $|\varphi(t)| \le (\|t\|^m + 1)/a(t)$, $x_*(t) \le e^{c(\|t\|^m+1)/a(t)}$ for a constant $c = c(\lambda^*)$, and (5). Then we may apply Fatou's lemma for a sequence $\varepsilon = \varepsilon_k \to 0$ to obtain, by (9) and (10), that

$$\mp \int_T a_j x_* d\mu = -\int_T \varphi x_* d\mu = \int_T \lim_{\varepsilon\to 0} F_\varepsilon d\mu \le \liminf_{\varepsilon\to 0} \int_T F_\varepsilon d\mu \le \mp g_j.$$

Hence $\int_T a_j x_* d\mu = g_j$. Since j was arbitrary in I, x_* is a solution of (7). The function $f_* := \rho x_*/a$ is then a solution of (2). By (4) and (6), $x_* \ln x_* \in L^1(T, \mu)$, i.e. $(af_*/\rho)\ln(af_*/\rho) \in L^1(T, \rho dt)$. Hence there are solutions f of (2) such that $H(f) > -\infty$. By the correspondence $f \leftrightarrow x = af/\rho$, the fact that f_* maximizes the functional H given at (b) is equivalent to saying that $\int_T x_* \ln x_* d\mu \le \int_T x \ln x\, d\mu$ for all the solutions $x \in L^1_+(T, \mu)$ of the problem (7). By Lemma 1 and Remark 3, (c) it suffices to show that $\int_T x_* \ln x_* d\mu \le \int_T x \ln x\, d\mu$ for any solution $x \in L^\infty_+(T)$ of (7). This holds by

$$\int_T x \ln x\, d\mu \ge P = D = \sum \lambda_i^* g_i - \int_T e^{\sum_i \lambda_i^* a_i - 1} d\mu = \int_T x_* \ln x_* d\mu.$$

The conclusion $P = D$ of Theorem 2 provides $-H(f_*) = L(\lambda^*)$. The uniqueness of λ^* and f_* (or, equivalently, x_*) follow from the strict convexity of $-L$, resp. $-H_\mu$ and the fact that T is not negligible, whence $p|_T = 0$ a.e. $\Rightarrow$ $p = 0$ for any polynomial

$p = \sum_{i\in I}\lambda_i X^i$ (the zeroes sets of nonconstant polynomials are algebraic varieties, and so have null Lebesgue measure). □

Proposition 5 develops an idea from Mead and Papanicolaou [32], that we generalize to our present context.

Proposition 5 *Let T, I, g and ρ, a satisfy the hypotheses of Theorem 4. Suppose also that $a(t) = \sum_{i\in I} c_i t^i$ and $\sum_{i\in I} c_i g_i > 0$. If $\sup L_{\rho,a,g} < \infty$, then there is a λ^* on which the supremum is attained, $\sup L_{\rho,a,g} = L_{\rho,a,g}(\lambda^*)$.*

Proof Since a is independent of $(t^i)_{i\in I\setminus\{0\}}$, $c_0 \neq 0$. Set $c_{i0} = c_i$, $c_{ij} = \delta_{ij}$ ($i \in I$, $j \in I\setminus\{0\}$). A change of variables $\lambda \mapsto \tilde\lambda$: $\lambda_i = \sum_{j\in I} c_{ij}\tilde\lambda_j$ gives $L(\lambda) = \tilde L(\tilde\lambda) := \sum_{j\in I}\tilde g_j\tilde\lambda_j - \int_T e^{\sum_{j\in I}\tilde\lambda_j\tilde a_j - 1}\rho\, dt$ where $\tilde g_j = \sum_{i\in I} c_{ij} g_i$ and $\tilde a_j = \tau_j/a$ for $\tau_j(t) = \sum_{i\in I} c_{ij}t^i$. Then $\sup\tilde L = \sup L$. We prove the attainment for $\tilde L$. Denote $\tilde\lambda$, $\tilde a_j$, $\tilde g$, $\tilde L$ by λ, a_j, g, L, respectively. Now $a_0 \equiv 1$, $g_0 > 0$ and $(\tau_i)_{i\in I}$ are linearly independent on any subset of positive measure. Let $\mu = \rho dt$. Since $\sup L < \infty$, $\mu(T) > 0$. Set $\lambda = (\lambda_0, \lambda')$ where $\lambda' = (\lambda_i)_{i\in I\setminus\{0\}}$. Maximizing L with respect to λ_0 gives $\alpha(\lambda') := -\ln\int_T e^{\sum_{i\in I\setminus\{0\}}\lambda_i a_i(t) - 1}d\mu(t)$ such that $\max_{\lambda_0} L(\lambda_0, \lambda') = L(\alpha(\lambda'), \lambda')$. Consider the (convex) potential $f(\lambda') := \int_T e^{\sum_{i\in I\setminus\{0\}}\lambda_i(a_i(t) - g_i)}d\mu(t)$ so that $\sup L < \infty \Leftrightarrow \inf f > 0$. If $\inf f$ is attained at some λ'_*, $\sup L$ will be attained at $(\alpha(\lambda'_*), \lambda'_*)$. By (3), $|\sum_{i\in I\setminus\{0\}}\lambda_i(a_i(t) - g_i)| \le \|\lambda'\|(c\frac{\|t\|^m+1}{a(t)} + \|g\|)$ where $\|\lambda'\| = \sum_{i\in I\setminus\{0\}}|\lambda'_i|$, $\|g\| = \max_{i\in I}|g_i|$ and c is a constant. Then for every sequence $\lambda'_k = (\lambda_{ki})_{i\in I\setminus\{0\}}$ such that $\lim_k \lambda'_k = \lambda'$ we have $e^{\sum_{i\in I\setminus\{0\}}\lambda_{ki}(a_i(t) - g_i)} \le e^{\sup_k\|\lambda'_k\|(c\frac{\|t\|^m+1}{a(t)} + \|g\|)} \in L^1(T, \mu)$. By Lebesgue's dominated convergence theorem, $\lim_k f(\lambda'_k) = f(\lambda')$. Thus f is continuous.

There is no $\lambda' \neq 0$ such that $p_{\lambda'}(t) := \sum_{i\in I\setminus\{0\}}\lambda_i(\tau_i(t)/a(t) - g_i) \le 0$ a.e. on T, for otherwise on the subset $Z : p_{\lambda'}(t) = 0$ of T we have $a(t)\sum_{i\in I\setminus\{0\}}\lambda_i g_i = \sum_{i\in I\setminus\{0\}}\lambda_i\tau_i(t)$; if $\sum_{i\in I\setminus\{0\}}\lambda_i g_i = 0$, we get $\mu(Z) = 0$ due to $\lambda' \neq 0$; if $\sum_{i\in I\setminus\{0\}}\lambda_i g_i \neq 0$, we get again $\mu(Z) = 0$ since a is independent of $(\tau_i)_{i\in I\setminus\{0\}}$ $(= (t^i)_{i\in I\setminus\{0\}})$. Hence Z is negligible. Then on $T\setminus Z$, $p_{\lambda'}(t) < 0$, $e^{rp_{\lambda'}(t)} \le 1$ $(r \ge 0)$ and by Lebesgue's theorem $f(r_k\lambda') = \int_T e^{r_k p_{\lambda'}(t)}d\mu(t) \to 0$ as $r_k \to \infty$, which is impossible since $\inf f > 0$. Then for any $\lambda' \neq 0$ there are a constant $\delta = \delta_{\lambda'} > 0$ and measurable subset $T_{\lambda'} \subset T$ with $\mu(T_{\lambda'}) > 0$ such that $p_{\lambda'}(t) \ge \delta$ for all $t \in T_{\lambda'}$. Hence $f(r\lambda') \ge \int_{T_{\lambda'}} e^{rp_{\lambda'}(t)}d\mu(t) \ge e^{r\delta}\mu(T_{\lambda'})$. Then for every $\lambda' \neq 0$, $\lim_{r\to\infty} f(r\lambda') = \infty$.

There is a compact $K \subset \mathbb{R}^{N-1}$ with $\inf f = \inf_K f$, for otherwise we could find a sequence of unit vectors λ'_k, and $r_k \to \infty$ such that $\lim_{k\to\infty} f(r_k\lambda'_k) = \inf f$; we can also assume there is a unit vector λ' such that $\lambda'_k \to \lambda'$. Given $r > 0$, $r\lambda'_k = s'\lambda'_k + (1 - s')r_k\lambda'_k$ for $s' = \frac{r_k - r}{r_k - 1}$ $(\to 1$ as $k \to \infty)$ whence $f(r\lambda'_k) \le s' f(\lambda'_k) + (1 - s')f(r_k\lambda'_k)$. Since $\sup_k|f(r_k\lambda'_k)| < \infty$ and f is continuous, letting $k \to \infty$ we get $f(r\lambda') \le f(\lambda')$ which is impossible because $\lim_{r\to\infty} f(r\lambda') = \infty$. Since $\inf f$ is attained on K, $\sup L$ will be attained. □

A more explicit outcome of Theorem 4 and Proposition 5 is the Corollary 6 from below, that for small ϵ is an approximate entropy maximization result.

Corollary 6 *Let $T \subset \mathbb{R}^n$ be a closed subset. Let $I \subset \mathbb{Z}_+^n$ be finite with $0 \in I$. Fix $k \in \mathbb{Z}_+$ such that $\max_{i \in I} |i| < 2k+2$. Let $(g_i)_{i \in I}$ be a set of numbers with $g_0 = 1$. Fix also an arbitrary constant $\epsilon > 0$. The following statements (a), (b), (c) are equivalent:*

(a) There exist functions $f \in L^1_+(T, dt)$ such that $\int_T |t^i| f(t) dt < \infty$ and

$$\int_T t^i f(t)\, dt = g_i, \quad i \in I; \tag{11}$$

(b) There exists a particular solution f_* of (11) maximizing the functional

$$H(f) = H_\epsilon(f) = -\int_T f \ln f\, dt - \epsilon \int_T \|t\|^{2k+2} f\, dt;$$

(c) The associated Lagrangian $L = L_\epsilon$ from below satisfies $\sup L < \infty$

$$L(\lambda) = \sum_{i \in I} g_i \lambda_i - \int_T e^{\sum_{i \in I} \lambda_i t^i - \epsilon \|t\|^{2k+2}} dt \; + 1.$$

In this case: $\sup L$ is attained in a point $\lambda^* = (\lambda_i^*)_{i \in I}$, both f_* and λ^* are uniquely determined, $-H(f_*) = L(\lambda^*)$ and

$$f_*(t) = e^{\sum_{i \in I} \lambda_i^* t^i - \epsilon \|t\|^{2k+2}},$$

in particular

$$\int_T t^i\, e^{\sum_{j \in I} \lambda_j^* t^j}\, e^{-\epsilon \|t\|^{2k+2}}\, dt = g_i \quad (i \in I).$$

Proof Use Theorem 4 for $a(t) \equiv 1$ and $\rho(t) = e^{-\epsilon \|t\|^{2k+2}}$ $(t \in T)$, which provides a Lagrangian $L_{\rho,a,g}$ and point $\lambda^*_{\rho,a,g}$ related to the present ones L, λ^* by $L_{\rho,a,g}(\lambda) = L(\lambda - \lambda^0)$ and $\lambda^* = \lambda^*_{\rho,a,g} - \lambda^0$ where $\lambda^0 = (\lambda_i^0)_{i \in I}$ with $\lambda_i^0 = \delta_{i0}$. Then Proposition 5 applies, since $\sum_{i \in I} c_i g_i = g_0 > 0$. □

Remarks Considering perturbations H_ϵ of the entropy H_0 as above might automatically provide enough control in the tails of a maximizing sequence to guarantee convergence by known arguments [33,36]. More specifically, a maximizing sequence converges in weak L^1, and if the dominant moments of order $2k+2$ were bounded, one could show that moments of lower order ≤2k converge. For this argument the author is indebted to the referee, who legitimately suggested that Corollary 6 and Theorem 4 may have shorter proofs by this standard method, and moreover results like Corollary 6 are rather known [18], see below. This seems to be true indeed, because maximizing $H_0 - \epsilon \int_T \|t\|^{2k+2} f(t) dt$ should provide a certain brake on the growth of the moments of order $2k+2$. However, within the quite technical hypotheses of Theorem 4 we could not find an obvious argument to get apriori bounds on the $2k+2$ moments, and for the sake of completeness we kept our initial proofs. Theorem 4 can at least unify and cover also other known cases, like $T :=$ compact, $\epsilon = 0$ [31], see also [2] (setting $a(t),\ \rho(t) \equiv 1$ on T) or, to some extent, (Theorem 8, [12]) setting $a(t) = (\|t\|^2+1)^k$, $\rho(t) = c\|t\|^{2-n}(\|t\|^2+1)^{-3/2}$ $(n \geq 2)$; we omit the details. Another application is for example Corollary 7 from below.

In principle, one could numerically maximize such Lagrangian functions L to obtain a vector λ^* and so a density f_*. Solving such dual problems (usually by Newton's method) turns to be the basic technique to this aim. The main effort is then to deal with the computational cost of approximating multiple integrals (like $\int_T t^j e^{\sum_{i\in I}\lambda_i t^i}\rho(t)dt$ if $a \equiv 1$, for instance) needed for the gradient of L [1,8,18,23,27–29].

Versions of Corollary 6 were indeed known, for instance in Groth and McDonald [18] that used an ansatz like $f_*(t) = e^{\sum_{i\in I}\lambda_i t^i}\rho(t)$ to derive a moment closure in the context of kinetic equations, when $t \in \mathbb{R}^3$ stands for the gas velocity [14]. Their density $\rho(t) = e^{\sigma(t)}$ with a certain negative $\sigma(t)$ of order $|\sigma(t)| \geq ct.\|t\|^{2k+2}$, see (Eqs. (51), (52), [18]) can be viewed as a "window" function that attenuates the distribution at high velocities. This type of modification to the maximum-entropy moment distribution has been proposed by Au [6], and Junk [21] and prevents the existence of very small packets of very fast particles that, as mentioned in [24], are the basic reason for non-solvability of the maximum entropy problem associated with Euler equations [14].

Our proposed technique has also similarities with the method used in Tagliani, [39] that deals with the particular case of the Hamburger moments problem for $T = \mathbb{R}$.

Corollary 7 *Let $T \subset \mathbb{R}^n$ be closed, $k \in \mathbb{Z}_+$, $I = \{i \in \mathbb{Z}_+^n : |i| \leq 2k\}$ and $(g_i)_{|i|\leq 2k}$ a set of reals with $g_0 = 1$. The statements (a), (b) are equivalent:*

(a) There exists an $f \in L^1_+(T, dt)$ such that

$$\int_T t^i f(t)\,dt = g_i \quad (|i| \leq 2k);$$

(b) $L(\lambda) := \sum_{|i|\leq 2k} g_i\lambda_i - \int_T e^{\frac{\sum_{|i|\leq 2k}\lambda_i t^i}{\|t\|^{2k}+1}} e^{-\|t\|^2-1}dt$ is bounded from above.

In this case, L attains its maximum in a point $\lambda^* = (\lambda_i^*)_{|i|\leq 2k}$ and

$$f_*(t) := \frac{1}{\|t\|^{2k}+1}\, e^{\sum_{|i|\leq 2k}\lambda_i^* \frac{t^i}{\|t\|^{2k}+1}}\, e^{-\|t\|^2-1}$$

satisfies $\int_T t^i f_*(t)\,dt = g_i$ $(|i| \leq 2k)$.

Proof Use Theorem 4 and Proposition 5 for $a = \|t\|^{2k}+1$, $\rho = e^{-\|t\|^2}$. □

Notation For $g = (g_i)_{|i|\leq 2k}$ having representing densities on $\mathbb{R}^n$, let $\lambda^* = \lambda_g^* = (\lambda_i^*)_{|i|\leq 2k}$ denote the vector maximizing $L_0(\lambda) = \sum_{|i|\leq 2k} g_i\lambda_i - \int e^{\sum_{|i|\leq 2k}\lambda_i t^i}dt$. Set

$p_g(t) = \sum_{|i|\le 2k} \lambda_i^* t^i$. Then $\sum_{|i|=2k} \lambda_i^* t^i \le 0$ for all $t \in \mathbb{R}^n$ (use $\int_{\mathbb{R}^n} e^{p_g} dt < \infty$ and polar coordinates). Let G be the set of all such g, with the property $\sum_{|i|=2k} \lambda_i^* t^i < 0$ for $t \neq 0$. Then (see [27]) G is dense and open in the set of all g having representing densities, consists of data g for which λ^* does provide a representing density $f_* = e^{p_g}$ of g maximizing $H(f) = -\int f \ln f dt$, and the map $G \ni g \mapsto \lambda_g^*$ is C^∞-diffeomorphic. Let n, $k = 2$, whence card$\{i : |i| \le 2k\} = 15$. Let $x = (x_i)_{i\in\mathbb{Z}_+^2,\, |i|\le 4}$ denote the variable in $\mathbb{R}^{15}$. Let $G_0 = \{g \in G : \det A\lambda^* \neq 0\}$ where $A = Ax$ is the matrix in (16). Then G_0 is dense and open in G. Given $g \in G$, we may set $g_j := \int t^j e^{p_g(t)} dt$ for $|j| \ge 5$.

Proposition 8 from below is reminiscent to Lasserre (Lemma 2, [28]), see also [29] or (Lemma 2, [26]), where similar recurrences were obtained. The idea in the case $n = 1$ is to compute integrals like $\int \frac{d}{dt}(t f_*(t))dt$ via Leibniz-Newton's formula. In our case $n = 2$ the basic idea is the same, but a careful application of Stokes' theorem will be required in the proof. Although such calculation is not a practical method itself for determining if a moment set comes from an underlying density, it could help to the approximation of λ^* when used along with suitable numerical techniques—see for example the use of Newton's method as in [28] together with the semidefinite programming methods for gradient and Hessian computation from [8,29].

Proposition 8 *Let $n, k = 2$ and $g \in G_0$. The higher order moments $(g_j)_{|j|\ge 5}$ of the maximum entropy density e^{p_g} can be expressed by relations of the form*

$$g_j = \sum_{|i|\le 4} r_{ji}(\lambda^*)\, g_i \qquad (j \in \mathbb{Z}_+^2,\ |j| \ge 5)$$

where $r_{ji} = r_{ji}(x)$ are universal rational functions, see (16)–(19).

Proof It suffices to prove that for any $j_0 \in \mathbb{Z}_+^2$ with $|j_0| \ge 5$ there are rational functions $c_{j_0 i} = c_{j_0 i}(\lambda^*)$, for $|i| < |j_0|$, such that $g_{j_0} = \sum_{|i|<|j_0|} c_{j_0 i} g_i$ and then proceed inductively. Set $|j_0| = l+1$ for $l \ge 4$ and denote $\lambda^* = (\lambda_i^*)_{|i|\le 4}$ by $x = (x_i)_{|i|\le 4}$. Set $x_\kappa = 0$ if $\kappa \not\ge 0$. Let $p = p_g$, namely $p(t) = \sum_{|\iota|\le 4} x_\iota t^\iota$. We will find a polynomial $\pi(t) = \sum_{|i|\le l} c_{j_0 i} t^i$ and a differential 1-form $\omega = e^p(u dt_1 + v dt_2)$ with u, v polynomials, depending on j_0, such that $d\omega(t) = (t^{j_0} - \pi(t))e^{p(t)} dt_1 \wedge dt_2$. By Stokes' theorem on disks D_r of center 0 and radius r, $\int_{D_r} d\omega = \int_{\partial D_r} \omega \to 0$ as $r \to \infty$ since ue^p, ve^p are rapidly decreasing ($g \in G$). Hence $\int_{\mathbb{R}^2} (t^{j_0} - \pi(t))e^{p(t)} dt_1 dt_2 = 0$ which is the desired conclusion. The condition on ω means that $L = L(u, v) := v\partial_1 p - u\partial_2 p + \partial_1 v - \partial_2 u$ where $\partial_m = \partial/\partial t_m$ $(m = 1, 2)$ satisfies $L = t^{j_0} - \pi$. We let $u(t) = \sum_{|j|=l-2} a_j t^j$, $v(t) = \sum_{|j|=l-2} b_j t^j$ with $a_j = a_j(x), b_j = b_j(x)$ rational functions to be determined. Set $e_1 = (1, 0), e_2 = (0, 1)$. In degree $l + 1$, the equation $L = t^{j_0} - \pi$ gives $\sum_{|j|=l-2,\, |\iota|=4,\, \iota_1\ge 1} b_j \iota_1 x_\iota t^{j+\iota-e_1} - \sum_{|j|=l-2,\, |\iota|=4,\, \iota_2\ge 1} a_j \iota_2 x_\iota t^{j+\iota-e_2} = t^{j_0}$. Change the summation indices by $i = j + \iota - e_{1,2}$ and identify the coefficients of t^i with $i \ge 0$, $|i| = l + 1$. Then

$$\sum_{|\iota|=4,\ (e_1\le\iota\le i+e_1)} \iota_1 x_\iota b_{i+e_1-\iota} - \sum_{|\iota|=4,\ (e_2\le\iota\le i+e_2)} \iota_2 x_\iota a_{i+e_2-\iota} = \delta_{ij_0} \quad (|i| = l + 1) \qquad (12)$$

where δ_{ij_0} is Kronecker's symbol. The summation conditions in the brackets from above may be omitted, since the terms outside the respective ranges vanish formally due to either $\iota_{1,2}=0$, or $a_j, b_j, x_\kappa = 0$ whenever $j, \kappa \not\geq 0$. Once we have such u, v, π is determined from $L = t^{j_0} - \pi$ by gathering all terms of degree $\leq l$ in $-L$. We solve (12) in the Appendix, that provides also an algorithm for computing $c_{j_0 i}, r_{ji}$ via the formulas (16)–(19). □

Corollary 9 below is an attempt to solve a maximum entropy problem by means of a system of ordinary differential equations (13) without computing multiple integrals. However this is rather a theoretic result, since it requires an accurate solution of (13) - that is, small increments Δs and so, a large number $1/\Delta s$ of iterations.

Corollary 9 *Let $n, k = 2$ and $g, g_0 \in \mathbb{R}^{15}$ such that $sg + (1-s)g_0 \in G_0$ for all $s \in [0, 1]$, where g_0 has a known $\lambda^*_{g_0}$. Set $\Gamma_i(x, s) = sg_i + (1-s)(g_0)_i$ for $|i| \leq 4$ and $\Gamma_j(x, s) = \sum_{|i|\leq 4} r_{ji}(x)\,(sg_i + (1-s)(g_0)_i)$ for $|j| \geq 5$ where $x = (x_i)_{|i|\leq 4} \in \mathbb{R}^{15}$. The system of ordinary differential equations*

$$\sum_{|j|\leq 4} \Gamma_{i+j}\,(x(s), s)\,\frac{dx_j}{ds}\,(s) = g_i - (g_0)_i \quad (|i| \leq 4); \quad x(0) = \lambda^*_{g_0} \tag{13}$$

has a C^∞ solution $x = x(s)$, defined on a neighborhood of $[0, 1]$, the matrix $[\Gamma_{i+j}(x(s), s)]_{|i|,|j|\leq 4}$ is defined and invertible for all $s \in [0, 1]$, and we have $x(1) = \lambda^*_g$.

Proof Since G_0 is open, the point $g(s) := sg + (1-s)g_0$ is in G_0 (in particular, has representing densities) for every s in a neighborhood of $[0, 1]$. Set $g(s) = (g_i(s))_{|i|\leq 4}$. Since $g(s) \in G$, it has a $\lambda^* = \lambda^*_{g(s)}$ maximizing $L_{0,\,g(s)}$. Let $x(s) = \lambda^*_{g(s)}$. Write $x(s) = (x_\iota(s))_{|\iota|\leq 4}$. Then $p_{g(s)}(t) = \sum_{|\iota|\leq 4} x_\iota(s) t^\iota$. The H-maximization holds and $e^{p_{g(s)}}$ is a representing density for $g(s)$,

$$g_i(s) = \int_{\mathbb{R}^2} t^i e^{\sum_{|\iota|\leq 4} x_\iota(s) t^\iota} dt\,(|i| \leq 4). \tag{14}$$

Denote by $g(s)_j$ for $|j| \geq 5$ the moments of higher order of $e^{p_{g(s)}}$, namely $g(s)_j := \int t^j e^{p_{g(s)}(t)} dt$ $(|j| \geq 5)$. Since the map $G \ni \tilde{g} \mapsto \lambda^*_{\tilde{g}}$ is diffeomorphic, $x(\cdot)$ is smooth and so we may apply d/ds to the equalities (14), whence

$$g_i - (g_0)_i = \sum_{|j|\leq 4} \int t^{i+j} e^{\sum_{|\iota|\leq 4} x_\iota(s) t^\iota} \frac{dx_j}{ds}(s) = \sum_{|j|\leq 4} g(s)_{i+j} \frac{dx_j}{ds}(s);$$

of course $g(s)_i = g_i(s)$ if $|i| \leq 4$. Also Then we obtain the differential equations (13) on a neighborhood of $[0, 1]$. The denominators of $r_{ji}(x)$ do not vanish on the set $\{x(s) : 0 \leq s \leq 1\}$ and so $\Gamma_{i+j}(x(s), s)$ are defined for $0 \leq s \leq 1$. Each matrix $[\Gamma_{i+j}(x(s), s)]_{|i|,|j|\leq 4} = [\int t^{i+j} e^{p_{g(s)}(t)} dt]_{|i|,|j|\leq 4}$ is positive definite and so invertible. By (14), $g_i = \int t^i e^{\sum_{|\iota|\leq 4} x_\iota(1) t^\iota} dt$ $(|i| \leq 4)$. Due to the uniqueness of the critical point of the Lagrangian $L_{0,g}$ we derive $x(1) = \lambda^*_g$. □

Remark 10 Since $\Gamma_j(x(s), s) = g(s)_j$ for $|j| \geq 5$ where $g(s) = sg + (1-s)g_0$, all the entries of the matrix $\Gamma = [\Gamma_{i+j}(x(s), s)]_{|i|,|j|\leq 4}$ of the system (13): $\Gamma(x(s), s) \cdot \frac{dx}{ds}(s) = g - g_0$ are moments and can be computed inductively by linear recurrences $g(s)_{j_0} = \sum_{|i|\leq l} c_{j_0 i}(x(s))g(s)_i$ ($|j_0| = l+1$), see (18), using the concrete formulas (16), (17) of $c_{j_o i}$; the explicit formulas of r_{ji} from (19) are not needed to this aim. Moreover, for each l the calculations of g_{j_0} ($|j_0| = l+1$) are independent of each other. We may consider any $g_0 \in G_0$, for instance the set of moments up to the 4th order of $e^{-t_1^4 - t_2^4}$. Also fast inversion algorithms exist for such Hankel matrices Γ. Then for problems of reasonable size one can use numerical methods for systems of ordinary differential equations to obtain λ_g^* $(= x(1))$.

The author is indebted to one of the referees for Remark 11 from below.

Remark 11 Due to the accuracy needed for its solution the system (13) is not a very practical way of computing λ^*, comparing to the more efficient Newton's method or its many variants [8,23,27–29] for the dual problem of maximizing L. Actually, one could use (13), written in the form $dx/ds = H(x(s))(g - g_0)$, to iteratively find $x(1)$ as follows. Let the increment $\Delta s = 1$. Then a forward Euler solve of the ODE gives $x_1 = x(0) - H(x(0))(g - g_0)$. The value x_1 is an estimate of $x(1)$, but not exactly, so one can repeat the process with x_1 as the new initial condition and a new moment g_1 which is computed from x_1. Doing this k times gives $x_k = x_{k-1} - H(x_k)(g - g_k)$. If we identify H with the Hessian of the dual problem, then this is just Newton's method. When is well conditioned, a standard Newton method for the dual problem will take only a handful of such iterations.

Acknowledgments The present work was supported by the Czech grant IAA 100190903 GA AV (RVO: 67985840) and the Romanian grant CNCS UEFISCDI, no. PN-II-ID-PCE-2011-3-0119. I wish to express my thanks to professor Marian Fabian for drawing the results of the Fenchel duality theory to my attention. Also, I thank professor Mihai Putinar for several interesting suggestions and relevant references. The author is also indebted to the reviewers for their expert indications and reference to more recent works, that definitely improved the paper.

Appendix: The functions $r_{j\,i}$

We give an algorithm to recurrently compute r_{ji}, c_{ji}, in particular solve (12) to finish the proof of Proposition 8. Set $\delta_k = \delta_{(l+1-k,k)\, j_0}$ for $0 \leq k \leq l+1$. Let $\alpha_k = a_{(l-2-k,k)}$, $\beta_k = b_{(l-2-k,k)}$ for $0 \leq k \leq l-2$. Thus $\alpha_k, \beta_k = 0$ for $k < 0, k \geq l-1$. Also $x_\kappa = 0$ if $\kappa \ngeq 0$. Change the summation indices in (12) by $j = i + e_{1,2} - \iota$ (≥ 0). Then (12) becomes $\sum_{|j|=l-2,(j_2\leq i_2)}(i_1 - j_1 + 1)x_{i-j+e_1}b_j - \sum_{|j|=l-2,(j_2\leq i_2)}(i_2 - j_2 + 1)x_{i-j+e_2}a_j = \delta_{i j_0}$ where the (redundant) condition $j_2 \leq i_2$ follows from $j \leq i$, that comes from $\iota \geq e_{1,2}$. For every $i = (l+1-k, k)$ with $0 \leq k \leq l+1$, we have the equivalence $(j \geq 0, |j| = l-2, j_2 \leq i_2) \Leftrightarrow j = (l-2-p, p)$ for $0 \leq p \leq k$ and hence the $l+1$ equations in (12) become now, respectively,

$$\sum_{p=0}^{k}[(4+p-k)x_{(4+p-k,k-p)}\beta_p - (k-p+1)x_{(3+p-k,k-p+1)}\alpha_p] = \delta_k, \quad 0 \leq k \leq l+1. \tag{15}$$

If $l \geq 5$, let $\alpha_0, \ldots, \alpha_{l-5} = 0$ and define $\beta_0, \ldots, \beta_{l-5}$ inductively by$4x_{40}\beta_k = -\sum_{p=0}^{k-1}(4+p-k)x_{(4+p-k,\,k-p)}\beta_p + \delta_k$ $(0 \leq k \leq l-5)$ where $\sum_{\emptyset} := 0$. Note that $x_{40} < 0$ since $g \in G$. This fulfills (15) for $0 \leq k \leq l-5$. Last six equations in (12) $[(l-4 \leq k \leq l+1$ in (15)$]$ will provide α_k, β_k $(l-4 \leq k \leq l-2)$. If $l = 4$, skip this step and go directly to the linear 6×6 system [(in this case (16) for $y, z, w = 0$]. In any case, we let now $i = (l+1-k, k)$ for $0 \leq k \leq l+1$ in (12). We have $i + e_1 - \iota = (l+2-k-\iota_1, k-\iota_2)$ and $i + e_2 - \iota = (l+1-k-\iota_1, k-\iota_2+1)$. Last 6 equations in (12) become $\sum_{|\iota|=4}\iota_1 x_\iota \beta_{k-\iota_2} - \sum_{|\iota|=4}\iota_2 x_\iota \alpha_{k-\iota_2+1} = \delta_k$ $(l-4 \leq k \leq l+1)$, see below. The brackets ( ) border quantities already known in terms of $\beta_0, \ldots, \beta_{l-5}$. The markers $\lfloor\ \rceil$ border sums of terms that are null due to $\iota_{1,2} = 0$, $\alpha_k, \beta_k = 0$ $(k \geq l-1)$ or $\alpha_k = 0$ $(0 \leq k \leq l-5)$:

$$4x_{40}\beta_{l-4} + (3x_{31}\beta_{l-5} + 2x_{22}\beta_{l-6} + 1x_{13}\beta_{l-7} + 0x_{04}\beta_{l-8})$$
$$\lfloor -0x_{40}\alpha_{l-3}\rceil - 1x_{31}\alpha_{l-4}\lfloor -2x_{22}\alpha_{l-5} - 3x_{13}\alpha_{l-6} - 4x_{04}\alpha_{l-7}\rceil = \delta_{l-4}$$

$$4x_{40}\beta_{l-3} + 3x_{31}\beta_{l-4} + (2x_{22}\beta_{l-5} + 1x_{13}\beta_{l-6} + 0x_{04}\beta_{l-7})$$
$$\lfloor -0x_{40}\alpha_{l-2}\rceil - 1x_{31}\alpha_{l-3} - 2x_{22}\alpha_{l-4}\lfloor -3x_{13}\alpha_{l-5} - 4x_{04}\alpha_{l-6}\rceil = \delta_{l-3}$$

$$4x_{40}\beta_{l-2} + 3x_{31}\beta_{l-3} + 2x_{22}\beta_{l-4} + (1x_{13}\beta_{l-5} + 0x_{04}\beta_{l-6})$$
$$\lfloor -0x_{40}\alpha_{l-1}\rceil - 1x_{31}\alpha_{l-2} - 2x_{22}\alpha_{l-3} - 3x_{13}\alpha_{l-4}\lfloor -4x_{04}\alpha_{l-5}\rceil = \delta_{l-2}$$

$$\lfloor 4x_{40}\beta_{l-1} + \rceil\, 3x_{31}\beta_{l-2} + 2x_{22}\beta_{l-3} + 1x_{13}\beta_{l-4} + \lfloor 0x_{04}\beta_{l-5}\rceil$$
$$\lfloor -0x_{40}\alpha_l - 1x_{31}\alpha_{l-1}\rceil - 2x_{22}\alpha_{l-2} - 3x_{13}\alpha_{l-3} - 4x_{04}\alpha_{l-4} = \delta_{l-1}$$

$$\lfloor 4x_{40}\beta_l + 3x_{31}\beta_{l-1} + \rceil\, 2x_{22}\beta_{l-2} + 1x_{13}\beta_{l-3} + \lfloor 0x_{04}\beta_{l-4}\rceil$$
$$\lfloor -0x_{40}\alpha_{l+1} - 1x_{31}\alpha_l - 2x_{22}\alpha_{l-1}\rceil - 3x_{13}\alpha_{l-2} - 4x_{04}\alpha_{l-3} = \delta_l$$

$$\lfloor 4x_{40}\beta_{l+1} + 3x_{31}\beta_l + 2x_{22}\beta_{l-1} + \rceil\, 1x_{13}\beta_{l-2} + \lfloor 0x_{04}\beta_{l-3}\rceil$$
$$\lfloor -0x_{40}\alpha_{l+2} - 1x_{31}\alpha_{l+1} - 2x_{22}\alpha_l - 3x_{13}\alpha_{l-1}\rceil - 4x_{04}\alpha_{l-2} = \delta_{l+1}.$$

Set $y = -3x_{31}\beta_{l-5} - 2x_{22}\beta_{l-6} - x_{13}\beta_{l-7}$, $z = -2x_{22}\beta_{l-5} - x_{13}\beta_{l-6}$ and $w = -x_{13}\beta_{l-5}$.
We easily read from above that α_k, β_k for $k = l-4,\ l-3,\ l-2$ are given by

$$\begin{bmatrix} 4x_{40} & 0 & 0 & x_{31} & 0 & 0 \\ 3x_{31} & 4x_{40} & 0 & 2x_{22} & x_{31} & 0 \\ 2x_{22} & 3x_{31} & 4x_{40} & 3x_{13} & 2x_{22} & x_{31} \\ x_{13} & 2x_{22} & 3x_{31} & 4x_{04} & 3x_{13} & 2x_{22} \\ 0 & x_{13} & 2x_{22} & 0 & 4x_{04} & 3x_{13} \\ 0 & 0 & x_{13} & 0 & 0 & 4x_{04} \end{bmatrix} \begin{bmatrix} \beta_{l-4} \\ \beta_{l-3} \\ \beta_{l-2} \\ -\alpha_{l-4} \\ -\alpha_{l-3} \\ -\alpha_{l-2} \end{bmatrix} = \begin{bmatrix} y + \delta_{l-4} \\ z + \delta_{l-3} \\ w + \delta_{l-2} \\ \delta_{l-1} \\ \delta_l \\ \delta_{l+1} \end{bmatrix} \tag{16}$$

(note also that $g \in G_0$). We have a_j, b_j, and so u, v such that $\deg(L(u,v) - t^{j_0}) \le l$. Now $\pi = t^{j_0} - L$ is determined by summing the terms of degree $\le l$ in $-L$. For $m = 1, 2$ set $K_m = \{(j, \iota) : |j| = l - 2, |\iota| \le 3, \iota_m \ge 1\}$. Then

$$\pi = \sum_{(j,\iota)\in K_2} a_j \iota_2 x_\iota t^{j+\iota-e_2} - \sum_{(j,\iota)\in K_1} b_j \iota_1 x_\iota t^{j+\iota-e_1} + \sum_{|j|=l-2,\, j_2\ge 1} j_2 a_j t^{j-e_2} - \sum_{|j|=l-2,\, j_1\ge 1} j_1 b_j t^{j-e_1}.$$

For any $i \ge 0$ with $|i| \le l$, the coefficient of t^i in the sum Σ_{K_2} from above is $\sum_{(j,\iota)\in K_2(i)} a_j \iota_2 x_\iota$ where $K_2(i) = \{(j,\iota) \in K_2 : j + \iota - e_2 = i\}$. The map $K_2(i) \ni (j,\iota) \mapsto i - j$ is bijective onto $I_i := \{\kappa \ge 0 : \kappa \le i, |\kappa| = |i| + 2 - l\}$. Then we may use it to change the summation index by $\kappa = i - j$ and get the coefficient of t^i in Σ_{K_2} as $\sum_{\kappa\in I_i} (\kappa_2 + 1) a_{i-\kappa} x_{\kappa+e_2}$. Similarly, the coefficient of t^i in Σ_{K_1} is $\sum_{\kappa\in I_i} (\kappa_1 + 1) b_{i-\kappa} x_{\kappa+e_1}$. The coefficient $c_{j_0 i}$ ($=$a rational function $c_{l j_0 i}(x)$ of x, actually) of t^i in $\pi(t)$ is then

$$c_{j_0 i} = \sum_{\kappa\in I_i} [(\kappa_2 + 1) x_{\kappa+e_2} a_{i-\kappa} - (\kappa_1 + 1) x_{\kappa+e_1} b_{i-\kappa}] + d_{j_0 i} \quad (|i| \le l) \tag{17}$$

where $d_{j_0 i} = (i_2 + 1) a_{i+e_2} - (i_1 + 1) b_{i+e_1}$ if $|i| = l - 3$, and 0 otherwise. We have

$$g_{j_0} = \sum_{|i|\le l} c_{j_0 i}(x)\, g_i \quad (|j_0| = l + 1,\ l \ge 4). \tag{18}$$

Successive compositions of the mapping $(g_i)_{|i|\le l} \mapsto ((g_{j_0})_{|j_0|=l+1}, (g_i)_{|i|\le l}) = (g_i)_{|i|\le l+1}$ given by (18) for $l = 4, 5, \ldots$ provide us with $r_{ji}(x)$ such that

$$g_j = \sum_{|i|\le 4} r_{ji}(x) g_i \quad (|j| \ge 5). \tag{19}$$

Thus (16)–(19) provide c_{ji}, r_{ji}. Since $\det Ax \ne 0$ and $x_{40} = \sum_{|i|=4} x_i t^i|_{t=e_1} < 0$, the denominators of the rational functions r_{ji} do not vanish at $x = \lambda^*$. □

It would be interesting to generalize Proposition 8 to arbitrary n and k, for a class of simple domains T including $\mathbb{R}^n$, $[0,\infty)^n$ and get rid of assumptions like $g \in G_0$, G, for Lagrangians L_ϵ with $\epsilon > 0$. Also, numerical tests of systems like (13) could be tried.

References

1. Abramov, R.V.: The multidimensional maximum entropy moment problem: a review on numerical methods. Commun. Math. Sci **8**(2), 377–392 (2010)
2. Ambrozie, C.-G.: Maximum entropy and moment problems. Real Anal. Exch. **29**(2), 607–627 (2003/04)
3. Ambrozie, C.-G.: Truncated moment problems for representing densities and the Riesz–Haviland theorem. J. Operator Theory. arXiv:1111.6555 (to appear)

4. Ambrozie, C.-G.: Finding positive matrices subject to linear restrictions. Linear Algebra Appl. **426**:2–3, 716–728 (2007)
5. Akhiezer, N.I.: The classical moment problem. Hafner Publ. Co., New York (1965)
6. Au, J.D.: Lösung Nichtlinearer problems in der Erweiterten Thermodynamik. PhD thesis, Technische Universität Berlin (2001)
7. Bakonyi, M., Woerdeman, H.J.: Maximum entropy elements in the intersection of an affine space and the cone of positive definite matrices. SIAM J. Matrix Anal. Appl. **16**(2), 369–376 (1995)
8. Bertsimas, D., Doan, X.V., Lasserre, J.B.: Approximating integrals of multivariate exponentials: a moment approach. Oper. Res. Lett. **36**(2), 205–210 (2008)
9. Borwein, J.M.: Maximum entropy and feasibility methods for convex and nonconvex inverse problems, optimization. J. Math. Program. Oper. Res. **61**:1(2012).
10. Borwein, J.M., Lewis, A.S.: Duality relationships for entropy-like minimization problems. SIAM J. Cont. Optim. **29**(2), 325–338 (1991)
11. Borwein, J.M., Lewis, A.S: Partially finite convex programming. I. Quasi relative interiors and duality theory. Math. Program. **57**:1, Ser. B, 15–48 (1992)
12. Blekherman, G., Lasserre, J.B.: The truncated K-moment problem for closure of open sets. J. Funct. Anal. **263**(11), 3604–3616 (2012)
13. Van Campenhout, J.M., Cover, T.M.: Maximum entropy and conditional probability. IEEE Trans. Inf. Theory. IT-27, 483–489 (1981)
14. Cercignani, C. (ed.): Kinetic Theories and the Boltzmann Equation. Lecture Notes in Mathematics, vol. 1048. Springer-Verlag, Berlin (1984)
15. Curto, R.E., Fialkow, L.A.: An analogue of the Riesz–Haviland theorem for the truncated moment problem. J. Funct. Anal. **255**(10), 2709–2731 (2008)
16. Curto, R.E., Fialkow, L.A.: Solution of the truncated complex moment problem for flat data. Mem. Amer. Math. Soc. **119**(568), x+32 (1996)
17. Fuglede, B.: The multidimensional moment problem. Exp. Math. **1**, 47–65 (1983)
18. Groth, C.P., McDonald, J.G.: Towards physically realizable and hyperbolic moment closures for kinetic theory. Continuum Mech. Thermodyn. **21**, 467–493 (2009)
19. Haviland, E.K.: On the momentum problem for distributions in more than one dimension, I. Am. J. Math. **57**, 562–568 (1935)
20. Jaynes, E.T.: On the rationale of maximum entropy methods. Proc. IEEE. **70**, 939–952 (1982)
21. Junk, M.: Minimum relative entropy systems of the Boltzmann equation (Unpublished work)
22. Junk, M.: Domain of definition of Levermore's five-moment system. J. Stat. Phys. **93**, 1143–1167 (1998)
23. Junk, M.: Maximum entropy for reduced moment problems. Math. Models Methods Appl. Sci. **10**:7(2000), 1001–1025
24. Junk, M.: Maximum entropy moment problems and extended Euler equations. Transp. Transition Regimes (MN, 2000), 169–198, IMA Vol. Math. Appl., 135, Springer, New York (2004)
25. Prestel, A., Delzell, C.N.: Positive polynomials. From Hilbert's 17th problem to real algebra. Springer Monographs in Mathematics. Springer, Berlin (2001)
26. Putinar, M., Budišić, M.: Conditioning moments of singular measures for entropy optimization I. Indagationes Mathematicae **23**(4), 848–883 (2012)
27. Hauck, C.D., Levermore, C.D., Tits, A.L.: Convex duality and entropy-based moment closures; characterizing degenerate densities. SIAM J. Control Optim. **47**:4, 1977–2015 (2008)
28. Lasserre, J.B.: Semidefinite programming for gradient and Hessian computation in maximum entropy estimation. Proceedings of the IEEE CDC conference, December (2007)
29. Lasserre, J.B.: A semidefinite programming approach to the generalized problem of moments. Math. Program. **112**:1, Ser. B, 65–92 (2008)
30. Léonard, C.: Minimization of entropy functionals. J. Math. Anal. Appl. **346**:1, 183–204 (2008)
31. Lewis, A.S.: Consistency of moment systems. Can. J. Math. **47**, 995–1006 (1995)
32. Mead, L.R., Papanicolaou, N.: Maximum entropy and the problem of moments. J. Math. Phys. **25**(8), 2404–2417 (1984)
33. Moreau, J.J.: Sur la polaire d'une fonction semi-continue supérieurement, Comptes Rendus de l'Academie des Sciences, vol. 258 (1964)
34. Putinar, M., Scheiderer, C.: Multivariate moment problems: geometry and indeterminateness. Ann. Sc. Norm. Super. Pisa Cl. Sci. (5) **5**(2), 137–157 (2006)

35. Putinar, M., Vasilescu, F.-H.: Solving moment problems by dimensional extension. Ann. Math. (2) **149**(3), 1087–1107 (1999)
36. Rockafellar, R.T.: Extension of Fenchel's duality for convex functions. Duke Math. J. **33**, 81–89 (1966)
37. Shohat, J., Tamarkin, J.: The problem of moments. Math. Surv. I, Am. Math. Soc., Providence, RI (1943)
38. Cichoń, D., Stochel, J., Szafraniec, F.H.: Riesz–Haviland criterion for incomplete data. J. Math. Anal. Appl. **380**:1, 94–104 (2011)
39. Tagliani, A.: Maximum entropy in the Hamburger moments problem. J. Math. Phys **35**(9), 5087–5096 (1994)
40. Vasilescu, F.-H.: Moment problems for multi-sequences of operators. J. Math. Anal. Appl. **219**(2), 246–259 (1998)