Preprint, Institute of Mathematics, AS CR, Prague. 2010-11-19 (rev. 1, 2010-12-7)

An application of the stationary phase method to maximum entropy solutions of the multivariable moments problems

Călin-Grigore Ambrozie*

December 7, 2010

Abstract

We use Hörmander's results on the method of the stationary phase to elaborate a technique of obtaining systems of algebraic equations, that can help the computation of the parameters defining the maximum entropy representing density of a finite set of moments.

Keywords: maximum entropy, moments problem, positive representing density.

Mathematics Subject Classification: MSC 44A60, 49J99

1 Statement of the problem

Fix n, $m \geq 1$ and let $\mathbb{R}^n$ be the n-dimensional Euclidian space, endowed with the Lebesgue measure dt, where $t = (t_1, \dots, t_n)$ denotes the variable in $\mathbb{R}^n$.

Let $A = A_{n,m} = \{\alpha \in \mathbb{Z}_+^n : |\alpha| \leq 2m\}$, where $|\alpha| = \alpha_1 + \cdots + \alpha_n$ for any multiindex α. Given an arbitrary set $\gamma = (\gamma_\alpha)_\alpha$ of numbers γ_α ($\alpha \in A$), the truncated problem of moments under consideration here requires to establish if there are nonnegative, absolutely continuous measures $\mu = f\, dt \geq 0$ on $\mathbb{R}^n$ such that

$$\int t^\alpha f(t)\, dt = \gamma_\alpha \quad (\alpha \in A). \tag{1}$$

*Supported by grants IAA100190903 of GA AV, 201/09/0473 GA CR

Thus we consider absolutely continuous representing measures $f\,dt$, with nonnegative density f from $L^1(\mathbb{R}^n)$ – the space of all classes of Lebesgue measurable functions that Lebesgue integrable on $\mathbb{R}^n$. Set $a := \operatorname{card} A$.

In a previous work [] we characterized the existence of such representing densities by the solvability of the following system

$$\int_{\mathbb{R}^n} t^\alpha \mathrm{e}^{\sum_{\beta\in A} x_\beta t^\beta}\, dt \;=\; \gamma_\alpha \qquad (\alpha \in A) \tag{2}$$

of a equations with a unknowns x_α ($\alpha \in A$). Therefore if our problem (1) has any absolutely continuous solution $\mu = f\,dt$, then it will necessarily have also a solution of the form from above. The concrete form of (2) then should allow to study the existence of (or approximate) the vector $x = (x_\alpha)_{\alpha\in A} \in \mathbb{R}^a$, see for instance [**?**], [3] and [].

For powers moment problems, it is known [], [] that if there exists an integrable representing density of the form $f_* = \exp\left(\sum_{\alpha\in A} x_\alpha u_\alpha\right)$ on the whole space $\mathbb{R}^n$, then knowing a large set of its moments, namely all γ_α, $\alpha \in A + A$, provides the values of x_α ($\alpha \in A$) by solving a compatible and determined linear system (**??**). Note the following example. Let $n = 1$ and $\gamma_0,\ \gamma_1,\ \gamma_2 \in \mathbb{R}$. Set $u_\alpha(t) = t^\alpha$ ($\alpha = 0, 1, 2$). In this case one can use (2) to compute x_α by hand. Namely, assume that $f_*(t) := \exp\,(x_0 + x_1 t + x_2 t^2)$, $t \in \mathbb{R}$ is integrable and satisfies (2). Since $f_* \in L^1(\mathbb{R})$, then $x_2 < 0$. Hence by the Leibniz–Newton formula we have $\int f_*' dt = 0$ and $\int (t f_*(t))' dt = 0$, where f' denotes the derivative of f. It follows $x_1\gamma_0 + 2x_2\gamma_1 = 0$ and $\gamma_0 + x_1\gamma_1 + 2x_2\gamma_2 = 0$. Then $x_1 = \gamma_0\gamma_1 d^{-1}$, $x_2 = -\gamma_0^2 d^{-1}$ and $x_0 = \ln(\gamma_0/\int \exp\,(x_1 t + x_2 t^2) dt)$, where $d := \gamma_0\gamma_2 - \gamma_1^2$. Hence $f_*(t) = C \exp\,[-(t-s)^2/d\,]$ is a multiple of the Gauss distribution of mean $s = \gamma_1/2$ and dispersion d. Thus we get the well–known fact that the maximum entropy probability density of given mean and dispersion is the normal one, see [11] for instance. Similar computations providing x in terms of the known data γ_α, $\alpha \in A$ can be done also when $A = \{\alpha = (\alpha_1, \ldots, \alpha_n) \in \mathbb{Z}_+^n \mid \alpha_1 + \cdots + \alpha_n \le 2\}$ (this moment problem has been solved in [8] by different methods).

Namely, f_* maximizes the Boltzmann's integral $-\int f \ln f dm$ amongst all the absolutely continuous measures $\mu = fm \ge 0$ satisfying the equalities (1).

To briefly recall the significance of the maximum entropy solution [7], [11], [12], let $V : (\Omega, \mathcal{A}, P)\mathbb{R}ightarrow(T, m)$ be a random variable with values in T and absolutely continuous repartition $P \circ V^{-1} = \mu = fm$, where $(\Omega, \mathcal{A}, P)$ is a probability field. Let T be finite with $m :=$ the normalized cardinal measure. The average of the minimum amount of information necessary to determine the position of V in T proves then to be equal to Shannon's entropy

$$H(f) := -\int_{\Omega} \log_2 f(V(\omega))\, dP(\omega) \quad (= -\sum_{t\in T} f(t) \log_2 f(t)),$$

see for instance [11]. In general, if T is endowed with some arbitrary non-negative measure m, then the corresponding degree of randomness of V is measured by

$$H(V) := -\int_{\Omega} \ln f \circ V\, dP \quad (= -\int_T f \ln f\, dm).$$

Suppose that the repartition f of V is unknown, but we can find the mean values of some quantities u_α, $\alpha \in A$ depending on V. The available data on V are thus given by the knowledge of the numbers

$$\gamma_\alpha := \int_{\Omega} u_\alpha(V(\omega))\, dP(\omega) \quad (= \int_T u_\alpha(t) f(t)\, dm(t)) \quad (\alpha \in A).$$

The problem is now to choose the most reliable f by using all this (and only this) information. The repartition f_* of the highest degree of randomness allowed by the conditions (1) is then the natural choice for f, see for instance [11], [12] for details. Note also in this sense the very interesting result from below.

Theorem 0 [7] *Let $n := 1$ and $T := [a, b] \subset \mathbb{R}$. Let V be a random variable with uniform distribution on T. If V_1, $V_2, \ldots$ are independent copies of V, then the conditional probability of V given the observation*

$$k^{-1} \sum_{i=1}^{k} u_\alpha(V_i) = \gamma_\alpha \qquad (\alpha \in A,\ k = 1, 2, \ldots)$$

converges to $f_{*,x}$ as $k\mathbb{R}ightarrow\infty$.

Therefore in certain moment–type problems it could be of interest to approximate $f_{*,x}$ (that is, $x \in \mathbb{R}^a$).

The main concern of the present paper is then to find a way of **computing / approximating the vector** $x = (x_\alpha)_\alpha$ **in the equation (2) from above**.

2 Main results

Let p be a polynomial of degree $2m$ in n variables $t = (t_1, \dots, t_n)$, with real coefficients x_i,

$$p(t) = \sum_{i \in \mathbb{Z}_+^n, |i| \le 2m} x_i t^i,$$

s.t. $p(t) \le -c\|t\|^2 + c'$ for all $t \in \mathbb{R}^n$, where c, $c' > 0$.

Set $x = (x_i)_i \in \mathbb{R}^N$, where $N := \operatorname{card}\{i : |i| \le 2m\}$.

Let $g_i = g_i(x)$ be defined by

$$g_i = \int_{\mathbb{R}^n} t^i e^{p(t)} dt \qquad (|i| \le 2m)$$

and set $g = (g_i)_i \in \mathbb{R}^N$. Thus $g = g(x)$.

Our problem is then to find a suitable way (analytic, numerical etc) of expressing x in terms of g; $x = x(g) =?$

Our **Main theorem** is the following.

Theorem *There exist $N-1$ nontrivial polynomial functions f_k of $N-1$ variables, the coefficients of which depend on g, s.t. the sets $\tilde{x} := (x_i)_{i \ne 0}$ satisfy*

$$f_1(\tilde{x}) = 0, \dots, f_{N-1}(\tilde{x}) = 0.$$

Lemma 1 *Let $C \subset \mathbb{R}^n$ be a closed convex cone and L, $M \subset \mathbb{R}^n$ be linear subspaces with $L \subset M$ and $\dim M/L = 1$ s.t. $L + C \cap M \ne M$. Let f be a linear functional on L s.t. $fx > 0$ for every nonzero $x \in C \cap L$. Then there exists a linear extension F of f to M s.t. $Fx > 0$ for every nonzero $x \in C \cap M$.*

Proof. We can suppose that $C \cap M \not\subset L$ (in particular, $C \cap M \ne \emptyset$). Fix also a unit vector $u \in M$, orthogonal to L. By a compactness argument, there is a constant $a > 0$ s.t.

$$d(x, C) \ge a\|x\| \qquad (x \in L,\ fx \le 0), \tag{3}$$

for otherwise we can find a sequence of unit vectors $x_k \in L$ with $fx_k \leq 0$ s.t. $d(x_k, C) \to 0$ as $k \to \infty$, and hence, a subsequence convergent to a unit vector $x \in C \cap L$ with $fx \leq 0$, contrary to the hypotheses.

Let $\mathcal{C} := \mathrm{ri}\,(C \cap M)$. We prove that $\mathcal{C} \cap L = \emptyset$. Suppose there exists a vector $v \in \mathcal{C}$ with $v \in L$. Let $c_1 \in (C \cap M) \setminus L$. Then the inner product $\langle c_1, u \rangle \neq 0$. Since v is in the relative interior $\mathcal{C}$ of the set $C \cap M$ and $c_1 \in C \cap M$, by [Theorem II.6.4, [?]] we can find an $\epsilon > 0$ s.t. $c_2 := -\epsilon c_1 + (1+\epsilon)v$ is in $C \cap M$. Since $v \in L$ and $u \perp L$, we have $\langle c_2, u \rangle = -\epsilon \langle c_1, u \rangle$. The number $\langle c_2, u \rangle$ is then $\neq 0$ and has opposite sign to $\langle c_1, u \rangle$. Write $c_i = \langle c_i, u \rangle u + h_i$ where $h_i \in L$ for $i = 1, 2$. Then $\langle c_i, u \rangle u \in (C \cap M) + L$. It follows, due to the signs of the coefficients, that both u, $-u \in C \cap M + L$, and so $\mathbb{R} \cdot u \in C \cap M + L$, whence $M = \mathbb{R} \cdot u + L \subset C \cap M + L$, that is contrary to the hypotheses $L + C \cap M \neq M$.

Since $\mathcal{C} \cap L = \emptyset$, one of the half-spaces associated to the hyperplane L in M must contain $\mathcal{C}$ entirely, for if $\mathcal{C}$ contained points x and y in the two opposing half-spaces, some point of the line segment between x and y would be in L, that is impossible. The corresponding closed half-space of M must then contain the closure

$$\overline{\mathcal{C}} = \overline{\mathrm{ri}\,(C \cap M)} = \overline{C \cap M} = C \cap M.$$

Then there is a unit vector $x_0 \in M$, namely one of the vectors u or $-u$ orthogonal to L in M, s.t. $\langle c, x_0 \rangle \geq 0$ for all $c \in C \cap M$. Extend f by taking $Fx_0 > \|f\| a^{-1}$. Then for any $c \in C \cap M$, the orthogonal decomposition

$$c = \lambda x_0 + h \qquad (\lambda \in \mathbb{R},\, h \in L)$$

gives $0 \leq \langle c, x_0 \rangle = \lambda \|x_0\|^2 + 0 = \lambda$. To prove that $Fc \geq 0$ with strict inequality if $c \neq 0$, consider two cases.

If $fh \geq 0$, we obtain $Fc = \lambda F x_0 + fh \geq 0$, and $Fc \neq 0$ unless both λ, $fh = 0$ which means $c = h \in C \cap L$ and $fh = 0$ that implies $c = 0$ by our hypotheses.

If $fh < 0$, by (3) we have

$$|fh| \leq \|f\| \, \|h\| \leq \|f\| \, a^{-1} d(h, C) \leq \|f\| a^{-1} \|h - c\| \leq \|f\| a^{-1} \lambda,$$

whence $Fc = \lambda F x_0 + fh \geq (Fx_0 - \|f\| a^{-1})\lambda \geq 0$, with strict inequality because $Fc = 0$ only when $\lambda = 0$ in which case $c = h \in C \cap L \, \mathbb{R}ightarrow \, fh \geq 0$ that is impossible when $fh < 0$.

For any multiindex $i = (i_1, \ldots, i_n) \in \mathbb{Z}_+^n$ we write as usual $i! = i_1! \cdots i_n!$, $|i| = i_1 + \cdots + i_n$ and $x^i = x_1^{i_1} \cdots x_n^{i_n}$ for a variable $x = (x_1, \ldots, x_n)$. Also, $i \leq j$ means $i_1 \leq j_1$, ..., $i_n \leq j_n$. Let $\deg p$ denote the degree of a polynomial p. Let p_h denote the homogeneous part of maximal degree of p.

Let $GL(n)$, resp. $O(n)$ denote as usual the group of all invertible, resp. orthogonal linear maps on $\mathbb{R}^n$.

Remind that a *positive definite form* in n variables is a polynomial $p = \sum_{i,j=1}^n a_{ij} X_i X_j$ s.t. the $n \times n$ matrix $[a_{ij}]_{i.j=1}^n$ is positive definite, namely $\sum_{i,j=1}^n a_{ij} x_i x_j > 0$ for every vector $(x_i)_{i=1}^n \neq 0$ in $\mathbb{R}^n$ or, equivalently, s.t. $p(x) \geq c\|x\|^2$ for some constant $c = c_p > 0$ ($\Leftrightarrow \lim_{\|x\| \to \infty} p(x) = +\infty$, too).

Definition We call an arbitrary polynomial $p \in \mathbb{R}[X]$ *positive definite* if there exist constants $c > 0$ and R s.t.

$$p(x) \geq c\|x\|^2$$

for all $x \in \mathbb{R}^n$ with $\|x\| \geq R$, or, equivalently, if there exist $c > 0$, c' s.t.

$$p(x) + c' \geq c\|x\|^2 \quad \forall\, x \in \mathbb{R}^n,$$

condition that easily proves also to be equivalent to

$$\lim_{\|x\| \to \infty} p(x) = +\infty.$$

Let $P = P_n = \{\, p \in \mathbb{R}[X_1, \ldots, X_n] \,:\, p \text{ is positive definite } \}$.

Remark 2 (a) If $p = \sum_{i,j=1}^n a_{ij} X_i X_j + \sum_{i=1}^n b_i X_i + c$, then $p \in P_n \Leftrightarrow$ the form $\sum_{i,j=1}^n a_{ij} X_i X_j$ is positive definite.

(b) P_n is a convex cone, stable under multiplication.

(c) If $p \in P_n$, then for every $T \in GL(n)$, $x_0 \in \mathbb{R}^n$ and $c \in \mathbb{R}$ the polynomial $p(TX + x_0) + c$ also is in P_n.

(d) If $X = (X^1, \ldots, X^k)$ is a partition of the set $X = (X_1, \ldots, X_n)$ of variables and $p_j \in \mathbb{R}[X^j] \subset \mathbb{R}[X]$ is a positive definite form in $\mathbb{R}[X^j]$ for each $j = \overline{1,k}$ then $p_1 + \cdots + p_k \in P_n$.

(e) P_n is the minimal set containing all polynomials $p_1 + \cdots + p_k$ with $1 \leq k \leq n$ from (e) and stable under the operations from (b) and (c).

(f) If $p \in P$, then $\deg p$ must be even ≥ 2.

(g) For p homegeneous, $p \in P \Leftrightarrow \inf_{\|x\|=1} p(x) > 0 \Leftrightarrow p(x) \geq c\|x\|^{deg\, p}\, \forall x$ for some $c > 0$.

(h) If the homogeneous part p_h of p is in P, then $p \in P$, but the converse is not true: for example, the polynomial $p = X_1^4 + X_2^2 \in \mathbb{R}[X_1, X_2]$ is in P_2 while $p_h = X_1^4 \notin P_2$.

We remind from [?] the following lemma.

Lemma 3 *For any $p \in \mathbb{R}[X]$ there exists a unique minimal linear subspace $Y \subset \mathbb{R}^n$ s.t. $p = p \circ P_Y$.*

Let $\operatorname{supp} p$ denote the unique minimal linear subspace provided by Lemma 3. We call $\operatorname{supp} p$ the *support* of the polynomial p.

Lemma 4 *Let $P : \mathbb{R}^n \to \mathbb{R}^n$ be linear s.t. $P^2 = P$ and $\dim im\, P = n - 1$. If $p \in \mathbb{R}[X]$ s.t. $p = p \circ P$, then $p = p \circ P_{ker(I-P^*)}$.*

Proof. Let $Z = \ker(I - P^*)$. Since P is a projection onto a hyperplane, $I - P$ is a projection onto a 1-dimensional space. Then there exist some vectors $v, w \in \mathbb{R}^n$ s.t. $x - Px = \langle x, v\rangle w$ for all $x \in \mathbb{R}^n$. The equality $P^2 = P$ is equivalent to $\langle v, w\rangle = 1$. We can assume that $\|w\| = 1$, replacing w by $\|w\|^{-1} w$ and v by $\|w\| v$. Set $e_1 = (1, 0, \ldots, 0) \in \mathbb{R}^n$. Let $O \in O(n)$ s.t. $Oe_1 = w$. Let $Q = O^* P O$ and $q = p \circ O$. Since $p = p \circ P$, we have $q \circ Q = q$. Write $O^* v = (a_1, \ldots, a_n)$. The equalities $1 = \langle v, w\rangle = \langle O^* v, O^* w\rangle = \langle (a_1, \ldots, a_n), e_1\rangle = a_1$ show that $a_1 = 1$. It follows that $Qx = x - \langle Ox, v\rangle O^* w = x - \langle x, O^* v\rangle e_1$. Hence for every $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$, we have $\langle (x_1, x_2, \ldots, x_n), (1, a_2, \ldots, a_n)\rangle = x_1 + a_2 x_2 + \cdots a_n x_n$ and so

$$Qx = (x_1, x_2, \ldots, x_n) - \langle (x_1, x_2, \ldots, x_n), (1, a_2, \ldots, a_n)\rangle (1, 0, \ldots, 0)$$

$$= (-\sum_{j=2}^{n} a_j x_j, x_2, \ldots, x_n).$$

Then $\partial_1 Q = 0$, that is, the polynomial function $Q = Q(x)$ does not depend on the variable x_1. Hence

$$Q(x_1, x_2, \ldots, x_n) \equiv Q(0, x_2, \ldots, x_n). \tag{4}$$

Now $(I - P)^* = (\langle \cdot, v\rangle w)^* = \langle \cdot, w\rangle v$ and hence $Z = \ker(I - P^*) = w^\perp$. Then for every $x = (x_j)_{j=1}^n \in \mathbb{R}^n$ we have

$$P_{O^* Z} x = O^* P_{w^\perp} O x = O^* (I - P_{\mathbb{R}\cdot w}) O x =$$

$$O^*(Ox - \langle Ox, w\rangle w) = x - \langle x, O^*w\rangle O^*w$$

$$= x - \langle x, e_1\rangle e_1 = (x_1, x_2, \ldots, x_n) - (x_1, 0, \ldots, 0) = (0, x_2, \ldots, x_n).$$

Then, using (4) also, we obtain $q(P_{O^*Z}x) = q(0, x_2, \ldots, x_n) = q(x)$, namely $q \circ P_{O^*Z} = q$. Hence $p \circ O P_{O^*Z} O^* = p$. But $P_{O^*Z} = O^* P_Z O$, and so, $p \circ P_Z = p$.

Lemma 5 *Let $\tilde{\pi}$, $\tilde{q}$, $\tilde{r}$ be polynomials with $deg\,\tilde{r} < deg\,\tilde{q}(< deg\,\tilde{\pi}?)$ and $\tilde{q}$ homogeneous of degree k. Write $\tilde{q} = \sum_{j=0}^{k} P_j X_n^j$ with $P_j \in \mathbb{R}[X']$ homogeneogous of degree $k - j$. Suppose there is an index $j \in \{1, \ldots, k-1\}$ s.t. $P_j \not\equiv 0$. Suppose also that $\tilde{\pi} \in \mathbb{R}[X']$. Then $e^{\tilde{\pi}+\tilde{q}+\tilde{r}} \notin L^1$.*

Lemma 6 *Let π, q, $r \in \mathbb{R}[X]$ s.t. $deg\,r < deg\,q(< deg\,\pi?)$ and q is homogeneous. Let $Y \subset \mathbb{R}^n$ be a linear subspace s.t. $\pi = \pi \circ P_Y$. Suppose that* $\sup\{d(z, Y) : z \in supp\,q\, \|z\| = 1,\, q(z) \geq 0\} = 1$. *Then $e^{\pi+q+r} \notin L^1$.*

Remind that we have obtained in [1] the following theorem.

Theorem 7 *Let $p \in \mathbb{R}[X_1, \ldots, X_n]$ be arbitrary. Set $f(t) = e^{p(t)}$ for $t \in \mathbb{R}^n$. The following statements are equivalent:*
(a) *The function $f = e^p$ is Lebesgue integrable on $\mathbb{R}^n$.*
(b) *The polynomial $-p$ is positive definite in $\mathbb{R}[X_1, \ldots, X_n]$.*

The idea is to be used firstly can be described by the following elementary example.
Example: $n = 1$, $m = 1$

In this case, the equations of moments are:
$\int e^{x_0+x_1t+x_2t^2}dt = g_0,\ \int te^{x_0+x_1t+x_2t^2}dt = g_1,\ \int t^2 e^{x_0+x_1t+x_2t^2}dt = g_2$

$\Rightarrow\ x_1g_0 + 2x_2g_1 = 0,\ \ g_0 + x_1g_1 + 2x_2g_2 = 0$

$\Rightarrow x_1 = x_1(g)$, $x_2 = x_2(g)$ by solving the system of equations $f_1(x_1, x_2) = 0$, $f_2(x_1, x_2) = 0$ from above

(while x_0 can be obtained from $\int_{\mathbb{R}} e^{x_0+x_1t+x_2t^2}dt = g_0$)

Proof: Leibniz-Newton formula

$\int_{-\infty}^{\infty} \frac{d}{dt}(e^{x_0+x_1t+x_2t^2})dt = e^{x_0+x_1t+x_2t^2}|_{t=-\infty}^{t=+\infty} = 0$

$\Rightarrow \int_{-\infty}^{\infty}(x_1+2x_2t)e^{x_0+x_1t+x_2t^2}dt = 0$, that is,

$x_1g_0 + 2x_2g_1 = x_1\int e^{x_0+x_1t+x_2t^2}dt + 2x_2\int te^{x_0+x_1t+x_2t^2}dt = 0$

and we similarly use $\int_{-\infty}^{\infty} \frac{d}{dt}(te^{x_0+x_1t+x_2t^2})dt = 0$

2.1 Notions of multivariable moments problems

Fix n, $m \in \mathbb{N}$
Problem:

Characterize those sets $g = (g_i)_{i\in\mathbb{Z}_+^n,\,|i|\le 2m}$ of real numbers g_i that admit nonnegative representing measures on $\mathbb{R}^n$ with respect to the powers t^i ($|i| \le 2m$), that is,

$$\int_{\mathbb{R}^n} t^i d\mu(t) = g_i \qquad (i \in \mathbb{Z}_+^n,\ |i| \le 2m)$$

where we used the multiindex notation,
$i = (i_1,\dots,i_n) \quad |i| = i_1+\cdots+i_n$
$t = (t_1,\dots,t_n) \quad t^i = t_1^{i_1}\cdots t_n^{i_n}$

$\mu : Bor(\mathbb{R}^n) \to [0,\infty)$ measure
s.t. $t^i \in L^1(\mathbb{R}^n, m)$ $\forall\ i$ with $|i| \le 2m$

We call μ a *representing measure* for g
We call $\int t^i d\mu(t)$ the *moments* of μ
If $\mu = f\,dt$ with $f \in L^1(\mathbb{R}^n, dt)$, we call f a *representing density* for g
Example 1 $n = 1$, $m =$ arbitrary, $g = (g_i)_{i=0}^{2m}$

Theorem (Hamburger, Markov, Chebyshev,...) A set $g = (g_0, g_1, \dots, g_{2m})$ is a sequence of moments of some nontrivial representing density $f \ge 0$, that is,

$$\int_{-\infty}^{\infty} t^i f(t)dt = g_i \qquad\qquad (i = 0, \dots, 2m),$$

if and only the Hankel matrix

$$H_g := [g_{i+j}]_{i,\,j \leq m}$$

is positive definite, namely $\sum_{i,j=0}^{m} g_{i+j}\lambda_i\lambda_j > 0$ for all $(\lambda_0, \dots, \lambda_m) \neq 0$, or equivalently,

$$g_0 > 0,\; g_0 g_2 - g_1^2 > 0,\; \dots,\; \det H_g > 0\,.$$

Proof

– Riesz-Haviland's theorem: g is a set of moments $\Leftrightarrow$ the functional $L : X^i \mapsto g_i$ satisfies $Lp \geq 0$ for all polynomials $p \geq 0$ ($Lp = \int p d\mu$)

– On the real line, $p \geq 0 \Leftrightarrow p = \sum q^2 =$ sum of squares of polynomials $q = \sum_i \lambda_i X^i$

– $L(q^2) = L(\sum_{i,j} \lambda_i\lambda_j X^{i+j}) = \sum_{i,j} \lambda_i\lambda_j g_{i+j}$

In this case (real line), various numerical algorithms can provide approximate solutions $\mu = f\,dt$

Example 2 $\qquad m = 1$, $n =$ arbitrary, $g = (g_i)_{|i| \leq 2}$

Since any polynomial of degree 2 in several variables is a sum of squares, we obtain the (also, well known):

Theorem A set $g = (g_{i_1,\dots,i_n})_{i_1+\cdots+i_n \leq 2}$ has representing measures $\mu \geq 0$ on $\mathbb{R}^n \Leftrightarrow$

$$\sum_{i,j \in \mathbb{Z}_+^n;|i|,|j| \leq m} g_{i+j}\lambda_i\lambda_j \geq 0$$

for all $(\lambda_i)_{|i| \leq m}$.

In this case (moments of order 2), there exist elementary ways of finding solutions μ.

In the general case, for arbitrary n and m (≥ 2), no such characterizations or analytic solutions are known (there are positive polynomials that are not sums of squares).

We remind from [] the following basic result.

Theorem Let $g = (g_i)_{\in \mathbb{Z}_+^n, |i| \leq 2m}$ be a set of powers moments of a measure $\mu = f dt + \nu \geq 0$, with $f \in L^1(\mathbb{R}^n, dt) \setminus \{0\}$ and ν singular with respect to dt. Namely,

$$\int_{\mathbb{R}^n} t^i d\mu(t) = g_i \ (|i| \leq 2m).$$

Then there exist $x_i \in \mathbb{R}$ ($|i| \leq 2m$), uniquely determined by g, such that the polynomial

$$p(t) := \sum_{|j| \leq 2m} x_j t^j$$

satisfies $p(t) \leq -c\|t\|^2 + c'$ and

$$\int_{\mathbb{R}^n} t^i \exp\,(\sum_{|j| \leq 2m} x_j t^j) dt = g_i \quad (|i| \leq 2m).$$

2.2 On the maximum entropy principle

Let

$$V : (\Omega, \mathcal{A}, P) \to (T, m)$$

be a random variable with values in T and absolutely continuous repartition

$$P \circ V^{-1} = \mu = fm,$$

where $(\Omega, \mathcal{A}, P)$ is a probability field and T is a measurable space.

If $T =$ finite and $m :=$ the normalized cardinal measure:

Theorem (Shannon) The average of the minimum amount of information necessary to determine the position of V in T equals the *entropy* $H(f)$ of V,

$$H(f) := -\int_{\Omega} \log_2 f(V(\omega))\, dP(\omega) \quad = -\sum_{t \in T} f(t) \log_2 f(t).$$

In general, the degree of randomness of V is measured by

$$H(V) := -\int_{\Omega} \ln f \circ V \, dP \quad (= -\int_T f \ln f \, dm).$$

Suppose the repartition f of V is unknown but we can find the average values g_i of some quantities u_i depending on V.

The available data on V are thus given by the knowledge of the numbers

$$g_i := \int_{\Omega} u_i(V(\omega)) \, dP(\omega) \; = \int_T u_i(t) f(t) \, dm(t) \tag{5}$$

The problem is now to choose the most reliable f, by using all this, and only this information.

Solution: $f = f_*$, maximizing $H(\,\cdot\,)$ subject to eqs. (5)
Formula: $f_*(t) = \exp \sum_i x_i u_i(t)$
Other motivations for H:

– Let $T = \mathbb{R}$ and $m = dt$;
Boltzmann's integral formula for the physical entropy,

$$H(f) = -\int_{\mathbb{R}} f(t) \ln f(t) dt.$$

– **Theorem** (Van Campenhout; Cover) Let $T = [a, b]$ be endowed with $m = dt$. Let V be a random variable with uniform distribution on T. Let $V_1, V_2, \ldots$ be independent copies of V.

Then the conditional probability of V given the observation

$$k^{-1} \sum_{p=1}^{k} u_i(V_p) = g_i \qquad (\, p = 1, 2, \ldots)$$

converges to f_* as $k \to \infty$.

Suppose we look for a joint repartition

$$fm := P \circ (V_1, \ldots, V_n)^{-1}$$

of n random variables $V_1, \ldots, V_n$ with values in $\mathbb{R}$ by knowing only the average values

$$g_i = \int_\Omega V_1^{i_1} \cdots V_n^{i_n} dP = \int_{\mathbb{R}^n} t_1^{i_1} \cdots t_n^{i_n} f(t) dt$$

for all multiindices $i = (i_1, \ldots, i_n)$ with $|i| \leq 2m$.

Then let $T := \mathbb{R}^n$, $m = dt$, $u_i(t) = t^i$ and maximize

$$H(f) := -\int f \ln f dm$$

among all absolutely continuous measures $\mu = fm \geq 0$ having the prescribed moments

$$\int t^i f(t) dt = g_i \qquad (|i| \leq 2m)$$

Conclusion: $f_*(t) = \exp p(t),\ p(t) = \sum_{|i| \leq 2m} x_i t^i$
Problem: computation of the coefficients x_i

3 Method of the stationary phase

$$\mathcal{M} = \mathcal{M}_{n,m} := \{i \in \mathbb{Z}_+^n : |i| \leq m, i \neq 0\}$$

$$M = M_{n,m} := \operatorname{card} \mathcal{M}$$

$$\tau : \mathbb{R}^n \to \mathbb{R}^M, \qquad \tau(t) := (t^i)_{i \in \mathcal{M}}$$

Lemma There is a map

$$a : \{i \in \mathbb{Z}_+^n : |i| \leq 2m\} \to \{\alpha \in \mathbb{Z}_+^M : |\alpha| \leq 2\}$$

s.t.

$$t^i \equiv \tau(t)^{a(i)} \ \ \forall i$$

Instead of the variables $t_1, \ldots, t_n$, we introduce new variables $T_1, \ldots, T_M$, s.t.

the monomials t^i of order $|i| \leq 2m$
can be expressed as
monomials T^α with $\alpha = a(i)$ of order $|\alpha| \leq 2$,
by
$t^i = T^\alpha|_{T=\tau(t)}$
Example $n = 1$, $m = 2$ $\quad \tau(t) = (t, t^2)$

$\mathcal{M} = \{1, 2\},\ M = 2; \qquad \mathbb{R}^n = \{t\}_{t\in\mathbb{R}},\ \mathbb{R}^M = \{(T_1, T_2)\}_{T_1,T_2\in\mathbb{R}}$

The variables T_1, T_2 are: "$T_1 = t$", "$T_2 = t^2$"
(dependent, $T_2 = T_1^2$, when restricted to the image of τ:

$t^0 = 1 = (t, t^2)^{(0,0)}$
$t^1 = T_1 = (t, t^2)^{(1,0)}$
$t^2 = T_1^2 = (t, t^2)^{(2,0)}$
$t^3 = T_1 T_2 = (t, t^2)^{(1,1)=a(3)}$; here $t^3 = \tau(t)^{a(3)}$)
$t^4 = T_2^2 = (t, t^2)^{(0,2)}$

The equations of moments $\int_{\mathbb{R}^n} t^i e^{p(t)} dt = g_i$ become

$$\int_{\mathbb{R}^M} T^\alpha e^{P(T)} d\mu(T) = g_i$$

where:
$P(T)$ =polynomial of degree 2 s.t. $P|_{T=\tau(t)} = p(t)$;

μ is a singular measure of integration along the n-dimensional submanifold $\{\tau(t)\}_t$ of $\mathbb{R}^M$;
write $\int T^\alpha e^{P(T)} d\mu(T) = \langle \mu, T^\alpha e^{P(T)} \rangle = g_i$
$\psi(T) := e^{-\|T\|^2}$
$T = (T_1, \ldots, T_M) \in \mathbb{R}^M$ independent variables

$\psi_k(T) := c_k \psi(kT) = c_k e^{-k^2\|T\|^2}$
c_k constant s.t. $\int_{\mathbb{R}^M} \psi_k(T) dT = 1\ \forall\ k \geq 1$

$\psi_k \to \delta$
in $\mathcal{D}'(\mathbb{R}^M)$, as $k \to \infty$

$\mu * \psi_k \to \mu * \delta = \mu$

$$\langle \mu * \psi_k, T^\alpha e^{P(T)} \rangle \to \langle \mu, T^\alpha e^{P(T)} \rangle = g_i. \tag{6}$$

$$\langle \mu * \psi_k, T^\alpha e^{P(T)} \rangle = \int_{\mathbb{R}^n} \int_{\mathbb{R}^M} \psi_k(T - \tau(\lambda)) T^\alpha e^{P(T)} dT \, d\lambda \tag{7}$$
$$= \int_{\mathbb{R}^M} T^\alpha d\tilde{\mu}(T),$$

$$\tilde{\mu} = [\, c_k \int_{\mathbb{R}^n} e^{-k^2 \|T - \tau(\lambda)\|^2 + P(T)} d\lambda \,] dT$$

$\tilde{\mu}$ is a continuous integral of gaussian densities

(6), (7) $\Rightarrow$ for large k, we get a small perturbation of the moments equations

$$\int_{\mathbb{R}^M} T^\alpha d\tilde{\mu}(T) \approx g_i$$

for which "the coefficients of p in e^p are computable"

For every fixed $\lambda \in \mathbb{R}^n$ and $j \in \mathcal{M}$ ($\subset \mathbb{Z}^n_+$), by Stokes' formula on large spheres, we have:

$$\int_{\mathbb{R}^M} \frac{d}{dT_j} (c_k e^{-k^2 \|T - \tau(\lambda)\|^2} \cdot e^{P(T)}) dT = 0 \;\; \Rightarrow$$
$$-2 \int_{\mathbb{R}^M} k^2 c_k e^{-k^2 \|T - \tau(\lambda)\|^2} (T_j - \lambda^j) e^{P(T)} dT$$
$$+ \int_{\mathbb{R}^M} \psi_k(T - \tau(\lambda)) \frac{d}{dT_j} (e^{P(T)}) dT = 0$$

($\psi_k(T) = c_k e^{-k^2 \|T\|^2}$). After integration over $\mathbb{R}^n$:

2nd term $= \langle \mu * \psi_k, \frac{d}{dT_j}(e^{P(T)}) \rangle \to \langle \mu, \frac{d}{dT_j}(e^{P(T)}) \rangle =$ a linear combination of the coefficients x_i, with coefficients depending on known data g

1st term = rational expression in terms of integrals of the form

$$\int u(y) e^{ikf(y)} dy$$

where $y =$ either T or t, and f is complex-valued
(for ex. $f(y) = i\|y - \tau(\lambda)\|^2$)

Theorem (Hörmander,...) Let $f = f(y)$ be a complex valued C^∞ function in a neighborhood of 0 in $\mathbb{R}^m$ s.t.

$\operatorname{Im} f \geq 0$, $f(0) = 0$, $f'(0) = 0$, $\det f''(0) \neq 0$.

Then there is a compact neighborhood $K = K_f$ of 0 s.t. for every $u \in C_0^\infty(K)$ and $p \geq 1$ we have

$$|\int ue^{ikf}dy - R_k \cdot (L_0 u + \frac{1}{k}L_1 u + \frac{1}{k^2}L_2 u + \cdots + \frac{1}{k^{p-1}})|$$
$$\leq C_p \frac{1}{k^{p+\frac{m}{2}}} \tag{8}$$

where $R_k = (\det(kf''(0))/2\pi i)^{-1/2}$

and each L_j is a differential operator of order $2j$ acting on u at 0, given by

$$L_j u = \sum_{\nu-\mu=j} \sum_{2\nu\geq 3\mu} i^{-j} 2^{-\nu} \langle f''(0)D, D\rangle^\nu (g^\mu u)(0)/\mu!\nu!$$

where $D = (\frac{1}{i}\frac{\partial}{\partial y_1}, \ldots, \frac{\partial}{\partial y_m})$ and

$$g(y) = f(y) - f(0) - \langle f''(0)y, y\rangle/2.$$

Moreover, the coefficients of L_j are rational homogeneous functions of degree $-j$ in $f''(0), \ldots, f^{(2j+2)}(0)$ with denominator $(\det f''(0))^{3j}$. In every term the total number of derivatives of u and f'' is at most $2j$.

Also, each constant $C_p = C_p(f, u)$ is bounded "when f, f', u are controlled".

Example of use of (8): $p = 2$, $m = N$, $y = T$,

$f(y) = i\|y - \tau(\lambda)\|^2$; for simplicity, $\lambda := 0$

$u(y) = y^\alpha e^{P(y)}$ with $\alpha \neq 0$;

we multiply the equation

$$\int ue^{ikf}dy = R_k(L_0 u + \frac{1}{k}L_1 u + O(\frac{1}{k^2}))$$
$$= R_k(u(0) + \frac{1}{k}(\Delta u)(0) + O(\frac{1}{k^2})) = R_k(\frac{1}{k}\Delta u(0) + O(\frac{1}{k^2}))$$

by k, then divide the result by

$$\int e^{if} dy = R_k \cdot (1 + O(\frac{1}{k}))$$

and obtain that

$$\frac{k \int u e^{ikf} dy}{\int e^{ikf} dy} = \frac{\Delta u(0) + O(\frac{1}{k})}{1 + O(\frac{1}{k^2})} = \Delta u(0) + O(\frac{1}{k}),$$

that provides

$$k \int e^{-k\|T-\tau(\lambda)\|^2} T^\alpha e^{P(T)} dT = (\Delta u) \cdot \int \psi_k(T - \tau(\lambda)) e^{P(T)} dT$$

$$+ O(1/k) \rightarrow (\Delta u) \times \text{known data}$$

Integration with resp. to λ gives, since $u = T^\alpha e^{P(T)}$, a **1st term** = quadratic function of x, with coefficients depending on g

etc

Conclusions:

– larger p are necessary to deal with higher order moments $m = 3, 4, \ldots$;

– also, f is not always quadratic; may be given by the implicit function theorem;

– this method can be used, in principle, for arbitrary data n, m etc;

– the usefullness of the results for concrete moments problems would only occur by means of explicitely computing the functions $f_i(X)$ in the main Theorem; this seems to be a routine, but difficult task, to be completed in future papers.

References

[1] C. Ambrozie, Infinite dimensional Fenchel duality and truncated moment problems in several real variables, to appear.

[2] C. Ambrozie, *Representing densities of the multi-sequences of moments*, Hokkaido Math. J., ...

[3] C.-G. Ambrozie, *A note on moment problems*, Proc. Operator Theory Conf. 17(2000), 21–27, Theta Found., Timisoara.

[4] N.I. Akhiezer, *The classical moment problem*, Hafner Publ. Co., New York, 1965.

[5] A. Atzmon, *A moment problem for positive measures in the unit disc*, Pacific J. Math. 59(1975), 317–325.

[6] M. Bakonyi; L. Rodman; I.M. Spitkovsky; H.J. Woerdeman, *Positive extensions of matrix functions of two variables with support in an infinite band*, C.R. Acad. Sci., Paris, Sér. I 323:8(1996), 859–863.

[7] J.M. Van Campenhout; T.M. Cover, *Maximum entropy and conditional probability*, IEEE Trans. Inf. Theory IT–27(1981), 483–489.

[8] R.E. Curto; L.A. Fialkow, *Solution of the truncated complex moment problem for flat data*, Memoirs of the A.M.S. 1996.

[9] C. Foiaş; A.E. Frazho, *A commutant lifting approach to interpolation problems*, Operator Theory: Advances and Applications, Birkhäuser, 1990.

[10] J.–P. Gabardo, *Trigonometric moment problems for arbitrary finite subsets of* $\mathbb{Z}^n$, Trans. Am. Math. Soc. 350:11(1998), 4473–4498 .

[11] S. Guiaşu, *Information theory with applications*, Mc–Graw Hill, London, 1977.

[12] E.T. Jaynes, *On the rationale of maximum entropy methods*, Proc. IEEE 70(1982), 939–952.

[13] G. Polya; G. Szegö, *Problems and theorems in Analysis*, II, Springer–Verlag, Berlin, 1971.

[14] W. Rudin, *The extension problem for positive–definite functions*, Illinois J. Math. 7(1963), 532–539.

[15] F.–H. Vasilescu, *Moment problems for multi–sequences of operators*, J. Math. Anal. Appl. 219:2(1998), 246–259.

Institute of Mathematics AV CR
Zitna 25, 115 67 Prague 1
Czech Republic

ambrozie@math.cas.cz