Pure and Applied Functional Analysis
Volume 3, Number 2, 2018, 393–402

Yokohama Publishers
ISSN 2189-3764 ONLINE JOURNAL

ON LAGRANGE MULTIPLIERS IN CONVEX ENTROPY MINIMIZATION

CONSTANTIN ZĂLINESCU

Abstract. Based on a characterization of the optimality of a feasible solution of a convex entropy minimization problem, one shows that the feasible solutions obtained using formally the Lagrange multipliers method are optimal.

1. Introduction

A common procedure to find the solutions of an optimization problem with a finite number of equality constraints is using the Lagrange multipliers method (LMM). More precisely, having the function $f : E \subset X \to \mathbb{R}$ to be minimized (maximized) with the constraints $g_i(x) = b_i$, where $g_i : E \to \mathbb{R}$ ($i \in \overline{1,m}$), is to consider the Lagrangian $L : E \times \mathbb{R}^m \to \mathbb{R}$ defined by

$$L(x, \lambda) := f(x) + \sum\nolimits_{i=1}^{m} \lambda_i (g_i(x) - b_i),$$

and to find the critical points $(\overline{x}, \overline{\lambda}) \in E \times \mathbb{R}^m$, that is

$$\nabla_x L(\overline{x}, \overline{\lambda}) = 0, \quad \nabla_\lambda L(\overline{x}, \overline{\lambda}) = 0. \tag{1.1}$$

So, in order to envisage LMM one must have the possibility to speak about $\nabla_x L(\overline{x}, \overline{\lambda})$; hence X must be a normed vector space (or, more generally, a topological vector space), $\overline{x}$ must be in the (algebraic) interior of E, and the functions f and g_i must be at least Gâteaux differentiable at $\overline{x}$. Moreover, the existence of $\overline{\lambda} \in \mathbb{R}^m$ verifying the conditions in Eq. (1.1) is a necessary condition for the optimality of $\overline{x}$ under supplementary conditions on the data; for a precise statement see for example [7, Th. 9.3.1]. Problems appear when the set E has empty (algebraic) interior, situation in which the differentiability of f and g_i can not be considered (see [7, pp. 171, 172]); this is often the case when X is a function-space, as in entropy minimization (or maximization) problems. However, in many books and articles on entropy optimization LMM is used in a formal way. Borwein and Limber (see [3]) describe the main steps of the usual procedure for solving the entropy minimization problem (see also the survey [1]); they mention "We shall see that this is usually the solution but each step in the above derivation is suspect and many are wrong without certain assumptions." Pavon and Ferrante (see [10, Cor. 9.3]) establish a

2010 *Mathematics Subject Classification.* 90C25, 49K99, 94A17.
Key words and phrases. Entropy minimization, Lagrange multipliers, sufficient optimality condition.

sufficient condition for the optimality of the element obtained using LMM. However, examining their application of this result for establishing that "the Gaussian density $p_c(x) = (2\pi)^{-1/2} \exp\left[-\frac{1}{2}\frac{x^2}{\sigma^2}\right]$ has maximum entropy among densities with given mean and variance", we observe that [10, Cor. 9.3] is not adequate for solving this problem.

The aim of this note is to show that the solutions found using formally LMM are indeed optimal solutions for the entropy minimization problem

$$(EM) \qquad \text{minimize } \int_T \varphi(x(t))d\mu(t) \ s.t. \ \int_T \psi_i(t)x(t)d\mu(t) = b_i \ (i \in \overline{1,m}),$$

where $\varphi : \mathbb{R} \to \overline{\mathbb{R}}$ is a proper convex function, $(T, \mathcal{A}, \mu)$ is a measure space, $\psi_i : T \to \mathbb{R}$ $(i \in \overline{1,m})$ are measurable, and $x \in X$ with X a linear space of measurable functions. In fact, we provide a characterization of a solution $\overline{x}$ of (EM) from which we deduce easily that $\overline{x}$ obtained using LMM is indeed a solution of the problem.

Note that J.M. Borwein and some of his collaborators treated rigorously problem (EM) when $\mu(T) < \infty$ and the functions ψ_i are from $L_\infty(T, \mathcal{A}, \mu)$ in a series of papers.

2. Preliminaries

Let $(T, \mathcal{A}, \mu)$ be a measure space. Set

$$\mathcal{M} := \mathcal{M}(T, \mathcal{A}, \mu) := \left\{x : T \to \overline{\mathbb{R}} \mid x \text{ is measurable}\right\},$$

where $\overline{\mathbb{R}} := \mathbb{R} \cup \{-\infty, \infty\}$ (with $\infty := +\infty$), and

$$\mathcal{M}_0 := \{x \in \mathcal{M} \mid x(t) \in \mathbb{R} \text{ for a.e. } t \in T\}, \quad \mathcal{M}_0^+ := \{x \in \mathcal{M}_0 \mid x \geq 0 \text{ a.e.}\}.$$

As usual we consider as being equal two elements of $\mathcal{M}$ which coincide almost everywhere (a.e. for short). Recall that for every function $x \in \mathcal{M}$ with values in $\overline{\mathbb{R}}_+ := [0, \infty]$ there exists its integral $\int_T x d\mu \in \overline{\mathbb{R}}_+$; moreover, if $\int_T x d\mu < \infty$, then $x \in \mathcal{M}_0$.

In the sequel we use the conventions

$$\infty - \infty := +\infty + (-\infty) := -\infty + \infty := \infty, \quad 0 \cdot (\pm\infty) := (\pm\infty) \cdot 0 := 0.$$

With these conventions $\int_T x d\mu := \int_T x_+ d\mu - \int_T x_- d\mu$ makes sense for every $x \in \mathcal{M}$, where $\alpha_+ := \max\{\alpha, 0\}$ and $\alpha_- := (-\alpha)_+$ for $\alpha \in \overline{\mathbb{R}}$; moreover, $\int_T x d\mu < \infty$ if and only if $\int_T x_+ d\mu < \infty$ (in particular $x_+ \in \mathcal{M}_0$), and $\int_T x d\mu \in \mathbb{R}$ if and only if $\int_T x_+ d\mu < \infty$ and $\int_T x_- d\mu < \infty$ (in particular $x \in \mathcal{M}_0$). The class of those $x \in \mathcal{M}$ with $\int_T x d\mu \in \mathbb{R}$ is denoted, as usual, by $L_1(T, \mathcal{A}, \mu)$, or simply $L_1(T)$, or even L_1.

Lemma 2.1. *Let $x, y \in \mathcal{M}$. Then the following assertions hold:*

(a) *If $x \leq y$ then $\int_T x d\mu \leq \int_T y d\mu$.*

(b) *If $x, y \geq 0$ and either $\int_T x d\mu < \infty$, or $\int_T y d\mu < \infty$, then $\int_T (x - y) d\mu = \int_T x d\mu - \int_T y d\mu$.*

(c) *If $\int_T x d\mu < \infty$ and $\int_T y d\mu < \infty$ then $\int_T (x + y) d\mu = \int_T x d\mu + \int_T y d\mu$*

We omit the proof which is standard and uses our conventions.

Remark 2.2. Observe that the hypotheses in assertions (b) and (c) of Lemma 2.1 are essential. For example, taking $T := \mathbb{R}_+$ endowed with the Lebesgue measure and $x(t) := y(t) := t$ for $t \in T$ we have that $x, y \geq 0$ and $\int_T x d\mu = \int_T y d\mu = \infty$, while $0 = \int_T (x - y) d\mu \neq \int_T x d\mu - \int_T y d\mu = \infty$; taking T, μ, x as before and $z := -y$, we have that $\int_T x d\mu = \infty$, $\int_T z d\mu = -\infty < \infty$ and $0 = \int_T (x + z) d\mu \neq \int_T x d\mu + \int_T z d\mu = \infty$.

Consider $\varphi \in \Lambda(\mathbb{R})$ (i.e. φ is a proper convex function on $\mathbb{R}$) with $\operatorname{int}(\operatorname{dom}\varphi) \neq \emptyset$. Because $[\varphi \leq \alpha] := \{u \in \mathbb{R} \mid \varphi(u) \leq \alpha\}$ is an interval, it follows immediately that $\varphi \circ x \in \mathcal{M}$ for every $x \in \mathcal{M}$, where $\varphi(\pm\infty) := \infty$. (The notations and notions which are not explained are standard; see for example [15].)

Let us consider a linear space $X \subset \mathcal{M}_0$, and define

$$\phi : X \to \overline{\mathbb{R}}, \quad \phi(x) := \int_T \varphi \circ x d\mu. \tag{2.1}$$

Proposition 2.3. *Let $\varphi \in \Lambda(\mathbb{R})$ with $\operatorname{int}(\operatorname{dom}\varphi) \neq \emptyset$, and let ϕ be defined by (2.1). Then*

$$\operatorname{dom}\phi = \left\{x \in X \mid (\varphi \circ x)_+ \in L_1\right\} \subset \{x \in X \mid x(t) \in \operatorname{dom}\varphi \ \text{a.e.}\}$$

and ϕ is convex; in particular, $\operatorname{dom}\phi$ is convex. Moreover, if φ is strictly convex (on its domain) and ϕ is finite on the convex set $K \subset \operatorname{dom}\phi$, then $\phi + \iota_K$ is strictly convex, where $\iota_K(x) := 0$ for $x \in K$, $\iota_K(x) := \infty$ for $x \in X \setminus K$.

Proof. The equality follows from our convention $\infty - \infty := \infty$, while the inclusion is obvious. Take $x, y \in \operatorname{dom}\phi$ and $\lambda \in]0, 1[$. Since φ is convex,

$$\varphi \circ (\lambda x + (1 - \lambda) y) \leq \lambda \cdot (\varphi \circ x) + (1 - \lambda) \cdot (\varphi \circ y) \quad \text{a.e.} \tag{2.2}$$

From Lemma 2.1 (a) we obtain that

$$\begin{aligned} \phi(\lambda x + (1 - \lambda) y) &= \int_T \varphi \circ (\lambda x + (1 - \lambda) y) d\mu \\ &\leq \int_T \left[\lambda \cdot (\varphi \circ x) + (1 - \lambda) \cdot (\varphi \circ y)\right] d\mu. \end{aligned}$$

Since $\phi(x) = \int_T (\varphi \circ x) d\mu < \infty$ and $\phi(y) = \int_T (\varphi \circ y) d\mu < \infty$, using Lemma 2.1 (c) we get

$$\begin{aligned} \int_T \left[\lambda \cdot (\varphi \circ x) + (1 - \lambda) \cdot (\varphi \circ y)\right] d\mu &= \lambda \int_T (\varphi \circ x) d\mu + (1 - \lambda) \int_T (\varphi \circ y) d\mu \\ &= \lambda \phi(x) + (1 - \lambda) \phi(y), \end{aligned}$$

and so $\phi(\lambda x + (1 - \lambda) y) \leq \lambda \phi(x) + (1 - \lambda) \phi(y)$. Hence ϕ is convex.

Let K be a convex subset of $\operatorname{dom}\phi$ such that $\phi(K) \subset \mathbb{R}$. Assume that φ is strictly convex, that is $s, s' \in \operatorname{dom}\varphi$ with $s \neq s'$ and $\lambda \in]0, 1[$ imply $\varphi(\lambda s + (1 - \lambda) s') < \lambda \varphi(s) + (1 - \lambda) \varphi(s')$. Moreover, assume by contradiction that there exist $x, y \in K$ with $\mu(T_0) > 0$, where $T_0 := \{t \in T \mid x(t) \neq y(t)\}$, and $\lambda \in]0, 1[$ such that $\phi(\lambda x + (1 - \lambda) y) = \lambda \phi(x) + (1 - \lambda) \phi(y)$, or, equivalently $\int_T \varphi \circ (\lambda x + (1 - \lambda) y d\mu = \int_T \left[\lambda \cdot (\varphi \circ x) + (1 - \lambda) \cdot (\varphi \circ y)\right] d\mu$. Since $\varphi \circ (\lambda x + (1 - \lambda) y) \leq \lambda \cdot (\varphi \circ x) + (1 - \lambda) \cdot (\varphi \circ y)$ a.e. and $\varphi \circ (\lambda x + (1 - \lambda) y)$, $\lambda \cdot (\varphi \circ x)$ and $(1 - \lambda) \cdot (\varphi \circ y)$ are from L_1, it follows that $\varphi \circ (\lambda x + (1-\lambda)y) = \lambda \cdot (\varphi \circ x) + (1-\lambda)\cdot(\varphi \circ y)$ a.e., contradicting our assumption that $\mu(T_0) > 0$. □

As well known, if ϕ takes the value $-\infty$, then it takes the value $-\infty$ on on the relative algebraic interior of $\operatorname{dom}\phi$ denoted $\operatorname{icr}(\operatorname{dom}\phi)$; however, $\operatorname{icr}(\operatorname{dom}\phi)$ is empty in many cases of interest when X is an L_p space with $p \in [1,\infty[$.

Having in view the applications to entropy minimization problems, in the sequel we consider $\varphi \in \Gamma(\mathbb{R})$ (that is $\varphi \in \Lambda(\mathbb{R})$ and φ is lower semicontinuous) such that φ is strictly convex on $I := \operatorname{dom}\varphi$, $\operatorname{int} I \neq \emptyset$, and φ is derivable on $\operatorname{int} I$; this implies that the conjugate φ^* of φ [defined by $\varphi^*(u) = \sup_{v\in\mathbb{R}} (uv - \varphi(v))$] is derivable on $\operatorname{int}(\operatorname{dom}\varphi^*)$ which is nonempty. Moreover, if $a := \inf I \in \mathbb{R}$, then either $\varphi(a) = +\infty$ and $\lim_{u\to a+}\varphi'(u) = -\infty$, or $\varphi(a) \in \mathbb{R}$ and $\varphi'(a) := \varphi'_+(a) = \lim_{u\to a+}\varphi'(u)$ ($\in [-\infty,\infty[$). Similarly, if $b := \sup I \in \mathbb{R}$, then either $\varphi(b) = +\infty$ and $\lim_{u\to b-}\varphi'(u) = +\infty$, or $\varphi(b) \in \mathbb{R}$ and $\varphi'(b) := \varphi'_-(b) = \lim_{u\to b-}\varphi'(u)$ ($\in\ ]-\infty,\infty]$). Assuming that ϕ is proper, (as seen above) ϕ is strictly convex on $\operatorname{dom}\phi$.

Proposition 2.4. *Consider $\overline{x}, x \in \operatorname{dom}\phi$ with $\phi(\overline{x}) \in \mathbb{R}$. Then*

$$\begin{aligned} \phi'(\overline{x}, x-\overline{x}) &:= \lim_{s\to 0+} \frac{\phi(\overline{x} + s(x-\overline{x})) - \phi(\overline{x})}{s} \\ &= \int_T \varphi'(\overline{x}(t)) \cdot \big(x(t) - \overline{x}(t)\big) d\mu(t). \end{aligned} \tag{2.3}$$

Proof. Since $\overline{x}, x \in \operatorname{dom}\phi$ we have that $\overline{x}(t), x(t) \in \operatorname{dom}\varphi$ a.e.

Assume first that $\phi(x) \in \mathbb{R}$. Take $(s_n)_{n\geq 1} \subset\ ]0,1[$ a decreasing sequence with $s_n \to 0$. Set

$$\theta_n := \varphi\circ x - \varphi\circ\overline{x} - \frac{\varphi\circ(\overline{x} + s_n(x-\overline{x})) - \varphi\circ\overline{x}}{s_n};$$

then $0 \leq \theta_n \leq \theta_{n+1}$ a.e. on T. Moreover

$$\lim_{n\to\infty}\theta_n(t) = \varphi(x(t)) - \varphi(\overline{x}(t)) - \varphi'(\overline{x}(t))\cdot(x(t) - \overline{x}(t)) \in [0,+\infty] \text{ for a.e. } t \in T.$$

By Lebesgue's monotone convergence theorem (see [11, Th. 1.26]), $(\varphi'\circ\overline{x})\cdot(x-\overline{x}) \in \mathcal{M}$ and

$$\begin{aligned} \phi(x) - &\phi(\overline{x}) - \phi'(\overline{x}, x-\overline{x}) \\ &= \lim_{n\to\infty}\left[\phi(x) - \phi(\overline{x}) - \frac{\phi(\overline{x} + s_n(x-\overline{x})) - \phi(\overline{x})}{s_n}\right] \\ &= \lim_{n\to\infty}\int_T \theta_n d\mu \\ &= \int_T \left[\varphi(x(t)) - \varphi(\overline{x}(t)) - \varphi'(\overline{x}(t))\cdot(x(t)-\overline{x}(t))\right] d\mu(t) \in \overline{\mathbb{R}}_+. \end{aligned}$$

Since $\varphi\circ x, \varphi\circ\overline{x} \in L_1$, we get the existence of $\int_T \varphi'(\overline{x}(t))\cdot(x(t)-\overline{x}(t))d\mu(t)$ as an element of $[-\infty,+\infty[$, and so (2.3) holds.

Assume now that $\phi(x) = -\infty$. In this case we have that $(\varphi'\circ\overline{x})\cdot(x-\overline{x}) \in \mathcal{M}$, too. With $(s_n)_{n\geq 1}$ as above, for each $n \geq 1$ we have that

$$\varphi'(\overline{x}(t))\cdot(x(t)-\overline{x}(t)) \leq \frac{\varphi(\overline{x}(t) + s_n(x(t)-\overline{x}(t))) - \varphi(\overline{x}(t))}{s_n} \quad \text{for a.e. } t \in T.$$

Using Lemma 2.1 (a) we get

$$\int_T \varphi'(\overline{x}(t)) \cdot (x(t) - \overline{x}(t)) d\mu(t) \leq \frac{\phi(\overline{x} + s_n(x - \overline{x})) - \phi(\overline{x})}{s_n} \quad \forall n \geq 1,$$

and so

$$\int_T \varphi'(\overline{x}(t)) \cdot (x(t) - \overline{x}(t)) d\mu(t) \leq \phi'_+(\overline{x}, x - \overline{x}). \tag{2.4}$$

Of course, because $\phi(x) = -\infty$ we have that

$$\phi(\lambda x + (1-\lambda)\overline{x}) = \lambda\phi(x) + (1-\lambda)\phi(\overline{x}) = -\infty \quad \forall \lambda \in]0,1[,$$

and so $\phi'_+(\overline{x}, x - \overline{x}) = -\infty$. We get (2.3) using (2.4). □

3. The entropy minimization problem

Let us consider $\psi_1, \ldots, \psi_m \in \mathcal{M}_0$ and the linear mappings

$$\Psi_k : X_k \to \mathbb{R}, \quad \Psi_k(x) := \int_T x\psi_k d\mu \quad (k \in \overline{1,m}),$$

where the linear space X_k is defined by

$$X_k := \{x \in \mathcal{M}_0 \mid x\psi_k \in L_1\}.$$

Take also $X_k^0 := \ker \Psi_k := \{x \in X_k \mid \Psi_k(x) = 0\}$ and set

$$\widetilde{X} := \textstyle\bigcap_{k=1}^m X_k, \quad \widetilde{X}^0 := \bigcap_{k=1}^m X_k^0;$$

note that $\widetilde{X} = \{x \in \mathcal{M}_0 \mid x\widetilde{\psi} \in L_1\}$, where $\widetilde{\psi} = |\psi_1| + \cdots + |\psi_m|$.

The entropy minimization problem is

$$(P) \qquad \text{minimize } \phi(x) \text{ s.t. } x \in X \cap \widetilde{X} \text{ with } \Psi_k(x) = b_k \ \forall k \in \overline{1,m},$$

where $b := (b_1, \ldots, b_m) \in \mathbb{R}^m$ is a given element.

Set

$$F_b := \left\{x \in \widetilde{X} \mid \Psi_k(x) = b_k \ \forall k \in \overline{1,m}\right\}. \tag{3.1}$$

Of course, if $\overline{x} \in F_b$ then $F_b = \overline{x} + \widetilde{X}^0$; in particular, F_b is a convex set. Because φ is strictly convex, if $\phi + \iota_{F_b}$ is proper then $\phi + \iota_{F_b}$ is strictly convex, and so (P) has at most one solution. Said differently, if $\overline{x}$ is a solution of (P) with $\phi(\overline{x}) \in \mathbb{R}$, then $\overline{x}$ is the unique solution of (P).

It is known (at least for the Boltzmann–Shannon entropy) that when problem (P) has a feasible solution $\widetilde{x} \in \operatorname{dom}\phi$ such that $\widetilde{x}(t) \in \operatorname{int}(\operatorname{dom}\varphi)$ for a.e. $t \in T$ and $\varphi'(a) = -\infty$ (when $a = \inf(\operatorname{dom}\varphi) \in \mathbb{R}$), $\varphi'(b) = +\infty$ (when $b = \sup(\operatorname{dom}\varphi) \in \mathbb{R}$), if $\overline{x}$ is the optimal solution of (P) with $\phi(\overline{x}) \in \mathbb{R}$, then $\overline{x}(t) \in \operatorname{int}(\operatorname{dom}\varphi)$ for a.e. $t \in T$. Indeed, assume that $a \in \mathbb{R}$ and $\mu(T_a) > 0$, where $T_a := \{t \in T \mid \overline{x}(t) = a\}$. Since $0 \leq \phi'_+(\overline{x}, \widetilde{x} - \overline{x})$, from (2.3) we have that

$$0 \leq \int_T \varphi'(\overline{x}(t)) \cdot (\widetilde{x}(t) - \overline{x}(t)) d\mu(t) \leq \phi(\widetilde{x}) - \phi(\overline{x}) < \infty,$$

whence $\int_{T_a} \varphi'(a) \cdot (\widetilde{x}(t) - a) d\mu(t) \in \mathbb{R}$. This is a contradiction, because $\varphi'(a) = -\infty$, $\widetilde{x}(t) - a > 0$ for $t \in T_a$ and $\mu(T_a) > 0$. We get a similar contradiction when $b = \sup(\operatorname{dom}\varphi) \in \mathbb{R}$ and $T_b := \{t \in T \mid \overline{x}(t) = b\}$ has positive measure.

Proposition 3.1. *Let $\overline{x} \in X \cap F_b$ be such that $\phi(\overline{x}) \in \mathbb{R}$.*

(a) *$\overline{x}$ is a solution of problem (P) if and only if (one of) the following two equivalent conditions hold(s):*

$$\phi'(\overline{x}, x - \overline{x}) \geq 0 \quad \forall x \in F_b \cap \operatorname{dom} \phi, \tag{3.2}$$

$$\int_T \varphi'(\overline{x}(t)) \cdot u(t) d\mu(t) \geq 0 \quad \forall u \in K_{\overline{x}} := [\mathbb{R}_+(\operatorname{dom} \phi - \overline{x})] \cap \widetilde{X}^0. \tag{3.3}$$

(b) *If there exists $\alpha_1, \ldots, \alpha_m \in \mathbb{R}$ such that $\varphi' \circ \overline{x} = \alpha_1 \psi_1 + \cdots + \alpha_m \psi_m$, then $\overline{x}$ is optimal solution of problem (P).*

Proof. . (a) The fact that [$\overline{x}$ is a solution of (P) iff (3.2) holds] follows immediately from a known result (see [6, Th. 3.8], [10, Th. 9.2], [14, Prop. 4]). Indeed, from the inequality $\phi'(\overline{x}, x - \overline{x}) \leq \phi(x) - \phi(\overline{x})$ we get the implication $\Leftarrow$. Assume that $\overline{x}$ is solution of (P) and take $x \in F_b \cap \operatorname{dom} \phi$. Then $(1-s)\overline{x} + sx \in F_b \cap \operatorname{dom} \phi$, and so $\phi\left((1-s)\overline{x} + sx\right) \geq \phi(\overline{x})$ for $s \in]0,1[$. Hence $s^{-1}\left[\phi\left((1-s)\overline{x} + sx\right) - \phi(\overline{x})\right] \geq 0$, and (3.2) follows taking the limit for $s \to 0$.

Since $(F_b \cap \operatorname{dom} \phi) - \overline{x} = (\operatorname{dom} \phi - \overline{x}) \cap \widetilde{X}^0$, and using Proposition 2.4, relation (3.2) can be rewritten as

$$\int_T \varphi'(\overline{x}(t)) \cdot u(t) d\mu(t) \geq 0 \quad \forall u \in (\operatorname{dom} \phi - \overline{x}) \cap \widetilde{X}^0,$$

which, at its turn, is clearly equivalent to (3.3).

(b) Consider the linear space

$$Y_{\overline{x}} := \{u \in \mathcal{M}_0 \mid (\varphi' \circ \overline{x}) \cdot u \in L_1\}$$

and the linear operator

$$\Theta_{\overline{x}} : Y_{\overline{x}} \to \mathbb{R}, \quad \Theta_{\overline{x}}(u) := \int_T \varphi'(\overline{x}(t)) \cdot u(t) d\mu(t).$$

Since $\phi'(\overline{x}, x - \overline{x}) < \infty$ for every $x \in \operatorname{dom} \phi$, from assertion (a) we have that $\overline{x}$ is a solution of (P) if and only if $K_{\overline{x}} - K_{\overline{x}} \subset Y_{\overline{x}}$ and $\Theta_{\overline{x}}(u) \geq 0$ for every $u \in K_{\overline{x}}$.

A sufficient condition for (3.3) is

$$Y := X \cap \widetilde{X} \subset Y_{\overline{x}} \text{ and } \Theta_{\overline{x}}(u) = 0 \ \forall u \in Y^0 := X \cap \widetilde{X}^0. \tag{3.4}$$

Assuming that $Y \subset Y_{\overline{x}}$, condition (3.4) is equivalent by [12, Lem. 3.9] to the existence of $\alpha_1, \ldots, \alpha_m \in \mathbb{R}$ such that $\Theta_{\overline{x}}|_Y = \alpha_1 \Psi_1|_Y + \cdots + \alpha_m \Psi_m|_Y$, or, equivalently,

$$\int_T \left[\varphi'(\overline{x}(t)) - (\alpha_1 \psi_1(t) + \cdots + \alpha_m \psi_m(t))\right] \cdot u(t) d\mu(t) = 0 \ \forall u \in Y. \tag{3.5}$$

Observing that an obvious sufficient condition for (3.5) is

$$\varphi'(\overline{x}(t)) = \alpha_1 \psi_1(t) + \cdots + \alpha_m \psi_m(t) \text{ for a.e. } t \in T, \tag{3.6}$$

the proof is complete. □

It is worth observing that (3.5) and (3.6) are equivalent when μ is σ-finite and the condition

(H) $\quad \forall A \in \mathcal{A}$ with $\mu(A) \in \mathbb{P} :=]0, \infty[$, $\exists u \in \mathcal{M}_0 : u > 0$ a.e. and $u\chi_A \in Y$

holds. (χ_A is the characteristic function of A, that is $\chi_A(t) := 1$ for $t \in A$, $\chi_A(t) := 0$ for $t \in T \setminus A$.) Indeed, the next result holds.

Proposition 3.2. *Let Y verify condition (H) and let μ be σ-finite. Assume that $\overline{y} \in \mathcal{M}$ is such that $\int_T \overline{y} u d\mu = 0$ for every $u \in Y$. Then $\overline{y} = 0$ (a.e.).*

Proof. By contradiction, assume that $\mu([\overline{y} \neq 0]) > 0$. Setting $A_+ := [\overline{y} > 0]$, $A_- := [\overline{y} < 0]$, we have that $\mu(A_+) > 0$ or $\mu(A_-) > 0$. We may assume that $\mu(A_+) > 0$ (otherwise replace $\overline{y}$ by $-\overline{y}$. Because T is σ-finite, there exists $A \in \mathcal{A}$ with $A \subset A_+$ such that $\mu(A) \in \mathbb{P} :=]0, \infty[$. By our hypothesis, there exists $u \in \mathcal{M}_0$ such that $u > 0$ and $u\chi_A \in Y$. It follows that $\int_T \overline{y} \cdot (u\chi_A) d\mu = 0$. Because $\overline{y} \cdot (u\chi_A) \geq 0$, it follows that $\overline{y} \cdot (u\chi_A) = 0$ a.e., and so $\overline{y}\chi_A = 0$ a.e. This is a contradiction because $\overline{y}(t) > 0$ for every $t \in A$ and $\mu(A) > 0$. □

It is worth observing that condition (3.6) is exactly the one found using formally LMM. In fact Proposition 3.1 and its proof explain how one arrives rigorously at the sufficient optimality condition of $\overline{x} \in F_b$ with $\phi(\overline{x}) \in \mathbb{R}$ in (3.6). An alternative justification of this fact in the case of countable sums is done in [13] and applied in [16]; of course this can also be obtained using Proposition 3.1 (b) for $T := \mathbb{N}^*$ and μ the counting measure (that is $\mu(A) = \infty$ for $A \subset \mathbb{N}^*$ infinite and $\mu(A)$ equals the number of elements of A for A finite).

Proposition 3.1 (b) shows that there is no need to verify separately that the solutions found using LMM in convex or concave entropy optimization are effectively solutions of (P); this verification is done for example in [5, Th. 12.1.1] and [4, Ths. 3.2, 3.3]. Note the following remark from [5, p. 410]: "The approach using calculus only suggests the form of the density that maximizes the entropy. To prove that this is indeed the maximum, we can take the second variation."

As in [2, Cor. 1], in the case $\mu(T) < \infty$ and $\psi_i \in L_\infty(T)$ ($i \in \overline{1,m}$), at least for Boltzmann–Shannon entropy ($\varphi(u) := u \ln u$ for $u \geq 0$ with $0 \ln 0 := 0$, and $\varphi(u) := \infty$ for $u < 0$), for $X = L_1(T)$ and $b \in \operatorname{icr} \mathcal{D}$ the problem (P) has optimal solution (provided by LMM), where

$$\mathcal{D} := \{b \in \mathbb{R}^m \mid F_b \cap \operatorname{dom} \phi \neq \emptyset\} = \left\{b \in \mathbb{R}^m \mid \exists x \in \operatorname{dom} \phi,\ \forall i \in \overline{1,m} : \int_T x\psi_i d\mu = b_i\right\}, \tag{3.7}$$

F_b being defined in (3.1).

The situation is completely different in the general case. Let us consider the problem

$$(PG)_m \qquad \text{minimize } \int_{\mathbb{R}} x(t) \ln x(t) dt \text{ s.t. } \int_{\mathbb{R}} t^{k-1} x(t) dt = b_k\ \forall k \in \overline{1,m}$$

with $b = (b_1, \ldots, b_m) \in \mathbb{R}^m$; $(PG)_m$ is studied for example in [5, Ch. 12] for $b_1 = 1$, and in [10] for $m = 3$ and $b = (1, 0, \sigma^2)$.

With our previous notation, $T := \mathbb{R}$, $\mathcal{A}$ is the class of Lebesgue measurable subsets of $\mathbb{R}$, and μ is the Lebesgue measure. Of course, $(PG)_m$ is a particular case of problem (P) in which φ is the Boltzmann–Shannon entropy. Of course, $\varphi \in \Gamma(\mathbb{R})$, $\operatorname{dom} \varphi = [0, \infty[$, $\varphi'(u) = 1 + \ln u$ if $u \in \operatorname{int}(\operatorname{dom} \varphi) =]0, \infty[$, $\varphi'(0) :=$

$\lim_{u\to 0+}\varphi'(u)=-\infty$, φ is strictly convex on $\operatorname{dom}\varphi$. We take $X:=\mathcal{M}_0$ and ϕ defined in (2.1); then $\operatorname{dom}\phi\subset\mathcal{M}_0^+$. In the present case $\psi_k(t)=t^{k-1}$ for $k\in\overline{1,m}$, and so $X_k=\{x\in\mathcal{M}_0\mid x\psi_k\in L_1\}$; hence $X_1=L_1$.

Let $m=3$; using Hölder's inequality for $p=q=2$ and $x\psi_1$, $x\psi_3$ with $x\in\mathcal{M}_0$ we get

$$\begin{aligned}\int_{\mathbb{R}}|tx(t)|\,dt&=\int_{\mathbb{R}}\sqrt{|x(t)|}\cdot\sqrt{t^2\,|x(t)|}dt\\&\le\sqrt{\int_{\mathbb{R}}|x(t)|\,dt}\cdot\sqrt{\int_{\mathbb{R}}t^2\,|x(t)|\,dt};\end{aligned}\tag{3.8}$$

equality holds in (3.8) for $x\in\widetilde{X}$ if and only if $x=0$ a.e. Hence, if $x\in X_1\cap X_3$ then $x\in X_2$, and so

$$\widetilde{X}=X_1\cap X_2\cap X_3=X_1\cap X_3=\{x\in L_1\mid x\psi_3\in L_1\}.$$

Proposition 3.3. *Consider the problem* $(PG)_3$. *Then*

$$\mathcal{D}=\{(0,0,0)\}\cup\left\{b\in\mathbb{R}^3\mid b_1>0,\ b_3>0,\ |b_2|\le\sqrt{b_1b_3}\right\}.$$

Moreover, if $b=0$, *then* $F_b\cap\operatorname{dom}\phi=\{0\}$, *and so* $\overline{x}:=0$ *is the solution of* $(PG)_3$. *If* $b_1,b_3>0$ *and* $|b_2|<\sqrt{b_1b_3}$ *then the solution and the value of* $(PG)_3$ *are*

$$\overline{x}(t)=\frac{b_1^2}{\sqrt{2\pi(b_1b_3-b_2^2)}}e^{-\frac{1}{2}\frac{b_1^2}{b_1b_3-b_2^2}(t-b_2)^2}\quad(t\in\mathbb{R})\tag{3.9}$$

and

$$\phi(\overline{x})=b_1\ln\frac{b_1^2}{\sqrt{2\pi e(b_1b_3-b_2^2)}},\tag{3.10}$$

respectively. In particular, if $b=(1,0,\sigma^2)$ *with* $\sigma>0$, *then*

$$\overline{x}(t)=\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{t^2}{2\sigma^2}}\ (t\in\mathbb{R}),\quad\phi(\overline{x})=-\ln\sqrt{2\pi e\sigma^2}.\tag{3.11}$$

Proof. Consider φ, X, ϕ, ψ_k, X_k as above. Hence $\widetilde{X}=X_1\cap X_3$.

Assume that $F_b\cap\operatorname{dom}\phi\neq\emptyset$ and take $x\in F_b\cap\operatorname{dom}\phi$; hence $x\ge 0$. Because $x\psi_1$, $x\psi_3\ge 0$, it follows that $b_1,b_3\ge 0$. Moreover, from (3.8) we obtain that $|b_2|\le\sqrt{b_1b_3}$.

If $b_1=0$, then $x\psi_1=0$ a.e., and so $x=0$ a.e.; it follows that $b_2=b_3=0$ and $F_b\cap\operatorname{dom}\phi=\{0\}$. The same conclusion is got when $b_3=0$.

Let $b_1,b_3>0$ and assume that $|b_2|=\sqrt{b_1b_3}$. Then equality holds in (3.8), which implies the existence of $\alpha\in\mathbb{R}\setminus\{0\}$ such that $x\psi_1=\alpha x\psi_3$. Since $\{t\in\mathbb{R}\mid\psi_1(t)=\alpha\psi_3(t)\}$ is finite, it follows that $x=0$ a.e., which implies that $b_1=0$, a contradiction. Therefore, $|b_2|<\sqrt{b_1b_3}$. The fact that $b_1,b_3>0$ and $|b_2|<\sqrt{b_1b_3}$ imply that $F_b\cap\operatorname{dom}\phi\neq\emptyset$ follows from the fact that $\overline{x}$ defined in (3.9) is the optimal solution of $(PG)_3$, as proved below.

Assume that $b_1,b_3>0$ and $|b_2|<\sqrt{b_1b_3}$. The problem is to find (if possible) some $\overline{x}\in F_b\cap\operatorname{dom}\phi$ $(\subset X\cap\widetilde{X})$ such that (3.6) holds. Assuming that such an $\overline{x}$ exists, then $\overline{x}(t)=e^{c_0+c_1t+c_2t^2}$ for $t\in\mathbb{R}$ and some $c_0,c_1,c_2\in\mathbb{R}$. Since $\overline{x}\in X_1=L_1$, we have necessarily that $c_2<0$, and so $\overline{x}(t)=e^{-\frac{1}{2}\alpha(t-\beta)^2+\gamma}$ for some $\alpha,\beta,\gamma\in\mathbb{R}$

with $\alpha > 0$ and every $t \in \mathbb{R}$. Imposing $\overline{x}$ to belong to F_b, and using the known fact that $\int_{\mathbb{R}} e^{-\frac{1}{2}t^2} dt = \sqrt{2\pi}$, we get

$$\alpha = \frac{b_1^2}{b_1 b_3 - b_2^2}, \quad \beta = \frac{b_2}{b_1}, \quad \gamma = \ln \frac{b_1^2}{\sqrt{2\pi(b_1 b_3 - b_2^2)}}.$$

Hence $\overline{x}$ is the function defined in (3.9). Moreover,

$$\phi(\overline{x}) = \int_{\mathbb{R}} \overline{x}(t) \ln \overline{x}(t) dt = \int_{\mathbb{R}} \left(-\frac{1}{2}\alpha(t-\beta)^2 + \gamma \right) \overline{x}(t) dt.$$

Taking into account the constraints and the expressions of α, β, γ above, we get the formula for $\phi(\overline{x})$ from (3.10).

Moreover, in the general case, for probability densities with mean $m \in \mathbb{R}$ and variance σ^2 ($\sigma > 0$), one has $b_1 = 1$, $b_2 = m$ and $b_3 = \sigma^2 + 2mb_2 - m^2 b_1 = \sigma^2 + m^2$. From (3.9) and (3.10) we get

$$\overline{x}(t) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(t-m)^2}{2\sigma^2}} \ (t \in \mathbb{R}), \quad \phi(\overline{x}) = -\ln \sqrt{2\pi e \sigma^2},$$

which gives (3.11) when $m = 0$. □

As mentioned above, problem $(PG)_3$ is considered for $b = (1, 0, \sigma^2)$ in [10] and solved applying [10, Cor. 9.3]. There $X = L_1(\mathbb{R})$, whence $X^* = L_\infty(\mathbb{R})$, and

$$\mathcal{V} := \left\{ x \in X \mid \int_{\mathbb{R}} x(t) dt = \int_{\mathbb{R}} t x(t) dt = \int_{\mathbb{R}} t^2 x(t) dt = 0 \right\}.$$

It is not explained why [10, (9.6)] holds and which is the annihilator of $\mathcal{V}$ in order to take $\overline{x}$ of the form $t \mapsto Ce^{\vartheta_1 t + \vartheta_2 t^2}$.

Proposition 3.3 provides an example in which $\mu(T) = \infty$ and the problem (P) has optimal solutions for all $b \in \mathcal{D}$. In [2] it is presented a situation with $X = L_1(0, \infty)$ and φ the Boltzmann–Shannon entropy in which (P) has optimal solutions for all $b \in \operatorname{icr} \mathcal{D}$, as in the case $\mu(T) < \infty$ and $\psi_i \in L_\infty(T)$.

Problem (P) is considered in [13] and [16] for $T := \mathbb{N}^*$, μ the counting measure, and $\varphi(u) = u \ln u - u$ for $u \geq 0$, $\varphi(u) = \infty$ for $u < 0$. Practically, a complete study of (P) for $m = 1$ is given in [13, Prop. 3.3]; so, besides providing the value of problem (P) for $b \in \mathcal{D}$ [$\mathcal{D}$ being defined in (3.7)], when $\mathcal{D} \neq \{0\}$ it is shown that either (P) has optimal solution for each $b \in \mathcal{D}$, or $(\operatorname{int} \mathcal{D}) \setminus \{b \in \mathcal{D} \mid (P)$ has optimal solution$\}$ is nonempty. In [13, Prop. 3.4], for $m = 2$ one has an example in which $\mathcal{D} = \{(0,0)\} \cup \big(]0, \infty[\times \mathbb{R}\big)$ and for every $b_1 > 0$ there exists only one $b_2 \in \mathbb{R}$ for which (P) has optimal solution which (moreover) can be found using formally LMM. A complete solution of problem (P) for $m = 2$ and $\psi_1 \equiv 1$ is given in [16, Th. 4.1]; the conclusions are similar to those in [13, Prop. 3.3] presented above.

The study of problem (P) for arbitrary measure spaces is done by P. Marechal in [8, 9] using a duality approach in which the primal space is similar to $\widetilde{X}$.

References

[1] J. M. Borwein, *Maximum entropy and feasibility methods for convex and nonconvex inverse problems*, Optimization **61** (2012), 1–33.

[2] J. M. Borwein, R. Choksi and P. Marechal, *Probability distributions of assets inferred from option prices via the principle of maximum entropy*, SIAM J. Optim. **14** (2003), 464–478
[3] J. M. Borwein and M. A. Limber, *On entropy maximization via convex programming*, Preprint CORR 92-16, University of Waterloo, 1992.
[4] K. Conrad, *Probability distributions and maximum entropy*, Expository paper, www.math.uconn.edu/˜kconrad/blurbs/analysis/entropypost.pdf
[5] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed., John Wiley & Sons, Hoboken, NJ, 2006.
[6] J. Jahn, *Introduction to the Theory of Nonlinear Optimization*, 3rd ed., Springer, Berlin, 2007.
[7] D. Luenberger, *Optimization by Vector Space Methods*, JohnWiley & Sons, Inc., 1969.
[8] P. Marechal, *On the principle of maximum entropy on the mean as a methodology for the regularization of inverse problems*, Grigelionis, B. (ed.) et al., Probability theory and mathematical statistics. Proceedings of the 7th international Vilnius conference, Vilnius, Lithuania, August, 12–18, 1998. Vilnius: TEV. 1999, pp. 481–492.
[9] P. Marechal, *A note on entropy optimization*, in: Approximation, Optimization and Mathematical Economics, M. Lassonde (ed)., Physica-Verlag, Heidelberg, 2001, pp. 205–211.
[10] M. Pavon and A. Ferrante, *On the geometry of maximum entropy problems*, SIAM Review **55** (2013), 415–439.
[11] W. Rudin, *Real and Complex Analysis*, (3rd edition), McGraw-Hill, Inc., 1987.
[12] W. Rudin, *Functional Analysis* (2nd edition), McGraw-Hill, Inc., 1991.
[13] C. Vallée and C. Zălinescu, *Series of convex functions: subdifferential, conjugate and applications to entropy minimization*, J. Convex Anal. **23** (2016), 1137–1160.
[14] C. Zălinescu, *On Gwinner's paper "Results of Farkas type"*, Numer. Funct. Anal. Optim. **10** (1989), 199–210.
[15] C. Zălinescu, *Convex Analysis in General Vector Spaces*, World Scientific, River Edge, NJ, 2002.
[16] C. Zălinescu, *On the entropy minimization problem in Statistical Mechanics*, J. Math. Anal. Appl. **457** (2018), 1713–1729.

Manuscript received March 30 2017
revised July 26 2017

C. Zălinescu
Faculty of Mathematics, University Al. I. Cuza, Bd. Carol I, Nr. 11, 700506 Iaşi, Romania, and "O. Mayer" Institute of Mathematics of the Romanian Academy, Iaşi, Romania.
E-mail address: `zalinesc@uaic.ro`