On entropy optimization and Lagrange multipliers

Constantin Zălinescu
Octav Mayer Institute of Mathematics, Iași, Romania
zalinesc@uaic.ro

Haifa, November 2021

Motivation

A common procedure to find the solutions of the problem

$$\text{minimize} \quad f(x) \quad \text{s.t.} \quad g_i(x) = b_i \ \ \forall i \in \overline{1, m},$$

where $f, g_i : E \subset X \to \mathbb{R}$, is to consider the Lagrangian $L : E \times \mathbb{R}^m \to \mathbb{R}$ defined by

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

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

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

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 n.v.s. (or, more generally, a t.v.s.), $\overline{x}$ 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 (CPL) is a necessary condition for the optimality of $\overline{x}$ under supplementary conditions on the data; for a precise statement see for example [L69, Th. 9.3.1][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 [L69, 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.

---

[1] [L69] D. Luenberger: *Optimization by Vector Space Methods*, JohnWiley & Sons, Inc. (1969).

Borwein & Limber in [BL96][2] describe the main steps of the usual procedure for solving the entropy minimization problem; 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 & Ferrante in [PV13, Cor. 9.3][3] establish a 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 observed that [PV13, Cor. 9.3] is not adequate for solving this problem.

---

[2][BL96] J.M. Borwein, M. Limber: *On entropy maximization via convex programming*, Preprint, Simon Fraser University (1996).

[3][PF13] M. Pavon, A. Ferrante: *On the geometry of maximum entropy problems*, SIAM Review 55 (2013), 415–439.

Aim

The entropy minimization problem we have in view is

$$\begin{aligned}&\text{minimize } \int_T \varphi(x(t))d\mu(t)\\&\text{s.t. } \int_T \psi_k(t)x(t)d\mu(t) = b_k \ (k = 1, \ldots, m),\end{aligned} \tag{EM}$$

where $(T, \mathcal{A}, \mu)$ is a measure space with μ a positive σ-finite measure, ψ_k is measurable for each k, and $\varphi : \mathbb{R} \to \overline{\mathbb{R}}$ is a proper lsc convex function with $\operatorname{int}(\operatorname{dom} \varphi) \neq \emptyset$.

Our aim is to show that the solutions found using formally LMM are indeed optimal solutions for the entropy minimization problem.

In fact, we provide a characterization of a solution $\overline{x}$ of (EM) from which we deduce that $\overline{x}$ obtained using LMM is 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.

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

$$\mathcal{M} := \mathcal{M}(T, \mathcal{A}, \mu) := \left\{x : T \to \overline{\mathbb{R}} \mid x \text{ is measurable}\right\},$$
$$\mathcal{M}_0 := \{x \in \mathcal{M} \mid x(t) \in \mathbb{R} \text{ for a.e. } t \in T\},$$
$$\mathcal{M}_0^+ := \{x \in \mathcal{M}_0 \mid x \geq 0 \text{ a.e.}\},$$

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

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

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

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 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 $\int_T (x + y) d\mu = \int_T x d\mu + \int_T y d\mu$.

Consider $\varphi \in \Lambda(\mathbb{R})$ (that 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 $\varphi(\pm\infty) := \infty$.

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

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

Proposition 3

Let $\varphi \in \Lambda(\mathbb{R})$ and Φ as above. 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 $x \in X \setminus K$.

As well known, if Φ takes the value $-\infty$, then it takes the value $-\infty$ on $\operatorname{icr}(\operatorname{dom}\Phi)$; 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 lsc) 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 $\varphi^*(v) = \sup_{v\in\mathbb{R}} (uv - \varphi(u))$] 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, then (as seen above) Φ is strictly convex on $\operatorname{dom}\Phi$.

Proposition 4

Consider $\overline{x}, x \in \operatorname{dom} \Phi$ with $\Phi(\overline{x}) \in \mathbb{R}$. Then

$$\Phi'(\overline{x}, x-\overline{x}) := \lim_{s\to 0+} \frac{\Phi(\overline{x}+s(x-\overline{x}))-\Phi(\overline{x})}{s} = \int_T \varphi'(\overline{x}(t))\cdot\big(x(t)-\overline{x}(t)\big)\,d\mu(t). \tag{R4}$$

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, for a.e. $t \in T$,

$$\theta(t) := \lim_{n\to\infty} \theta_n(t) = \varphi(x(t)) - \varphi(\overline{x}(t)) - \varphi'(\overline{x}(t)) \cdot (x(t) - \overline{x}(t)) \geq 0$$

By Lebesgue's monotone convergence theorem (see [R87, Th. 1.26][4]), $\theta = (\varphi' \circ \overline{x}) \cdot (x - \overline{x}) \in \mathcal{M}$, and

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

Since $\varphi \circ x, \varphi \circ \overline{x} \in L_1$, we get the existence of $\int_T \varphi'(\overline{x}(t)) \cdot (x(t) - \overline{x}(t)) d\mu(t) \in [-\infty, +\infty[$, and so (R4) holds.
The case $\varphi(x) = -\infty$ is simpler (using Lemma 1).

[4] [R87] W. Rudin, *Real and Complex Analysis*, (3rd edition), McGraw-Hill, Inc., 1987.

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

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

note that $\widetilde{X} = \{x \in \mathcal{M}_0 \mid x\widetilde{\psi} \in L_1\}$, where $\widetilde{\psi} = |\psi_1| + \ldots + |\psi_m|$. The entropy minimization problem in this context is

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

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

Set

$$F_b := \{x \in \widetilde{X} \mid \Psi_k(x) = b_k \ \forall k \in \overline{1,m}\}. \tag{R6}$$

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. $t \in T$.

Proposition 5

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):

$$\Phi'(\overline{x}, x - \overline{x}) \geq 0 \quad \forall x \in F_b \cap \operatorname{dom} \Phi, \tag{R7}$$

$$\int_T \varphi'(\overline{x}(t)) \cdot u(t) d\mu(t) \geq 0 \quad \forall u \in K_{\overline{x}}, \tag{R8}$$

where $K_{\overline{x}} := [\mathbb{R}_+(\operatorname{dom} \Phi - \overline{x})] \cap \widetilde{X}^0$.

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

Proof. (a) The equivalence

$$[\overline{x} \text{ is a solution of (P)}] \Longleftrightarrow [\Phi'(\overline{x}, x - \overline{x}) \geq 0 \quad \forall x \in F_b \cap \operatorname{dom} \Phi]$$

follows immediately from a known result.
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 so, taking the limit for $s \to 0$, $\Phi'(\overline{x}, x - \overline{x}) \geq 0$.

Since $(F_b \cap \operatorname{dom} \Phi) - \overline{x} = (\operatorname{dom} \Phi - \overline{x}) \cap \widetilde{X}^0$, and using Proposition 4, (R7) can be rewritten as

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

which, at its turn, is clearly equivalent to (R8).

(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

$$\Theta_{\overline{x}} : Y_{\overline{x}} \to \mathbb{R}, \quad \Theta_{\overline{x}}(u) := \int_T (\varphi' \circ \overline{x}) \cdot u d\mu.$$

Since $\Phi'(\overline{x}, x - \overline{x}) < \infty$ for every $x \in \operatorname{dom} \Phi$, from assertion (a), $\overline{x}$ is solution of (P) $\iff$ $K_{\overline{x}} - K_{\overline{x}} \subset Y_{\overline{x}}$ and $\Theta_{\overline{x}}(u) \geq 0 \ \forall u \in K_{\overline{x}}$.
A sufficient condition for (R8) is

$$Y := X \cap \widetilde{X} \subset Y_{\overline{x}} \text{ and } \Theta_{\overline{x}}(u) = 0 \ \ \forall u \in Y^0 := X \cap \widetilde{X}^0. \quad \text{(R9)}$$

For $Y \subset Y_{\overline{x}}$, condition (R9) is equivalent to the existence of $\alpha_1, \ldots, \alpha_m \in \mathbb{R}$ such that $\Theta_{\overline{x}}|_Y = \alpha_1 \Psi_1|_Y + \ldots + \alpha_m \Psi_m|_Y$.

Clearly, $\Theta_{\overline{x}}|_Y = \alpha_1 \Psi_1|_Y + \ldots + \alpha_m \Psi_m|_Y$ is equivalent to

$$\int_T \left[\varphi'(\overline{x}(t)) - (\alpha_1 \psi_1(t) + \ldots + \alpha_m \psi_m(t))\right] \cdot u(t) d\mu(t) = 0 \quad \forall u \in Y. \tag{R10}$$

Observing that an obvious sufficient condition for (R10) is

$$\varphi'(\overline{x}(t)) = \alpha_1 \psi_1(t) + \ldots + \alpha_m \psi_m(t) \quad a.e. \ on \ T, \tag{R11}$$

the proof is complete. □

It is worth observing that (R10) and (R11) are equivalent when μ is σ-finite and the condition

(H) $\forall A \in \mathcal{A}$ with $\mu(A) \in]0, \infty[$, $\exists u \in \mathcal{M}_0$ such that $u > 0$ 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 $t \in T \setminus A$.)

It is worth observing that condition (R11) is exactly the one found using formally LMM. In fact Proposition 5 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 (R11).

An alternative justification of this fact in the case of countable sums is done by Vallée & Z. in [VZ16][5] and applied in [Z18][6]; of course this can also be obtained using Proposition 5 (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).

[5][VZ16] C. Vallée, C. Z.: *Series of convex functions: subdifferential, conjugate and applications to entropy minimization*, J. Convex Anal. 23(4) (2016), 1137-1160.

[6][Z18] C. Z.: *On the entropy minimization problem in Statistical Mechanics*, J. Math. Anal. Appl. 457 (2018), 1713-1729].

Proposition 5 (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 the proof of [CT06, Th. 12.1.1][7] and in the proofs of [C, Ths. 3.2, 3.3][8].

Note the following remark from [CT06, page 420]:

"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."

[7] [CT06] T. M. Cover, J. A. Thomas: *Elements of Information Theory*, 2nd ed., John Wiley & Sons, Hoboken, NJ, 2006.

[8] [C] K. Conrad: *Probability distributions and maximum entropy*, Expository paper, www.math.uconn.edu/~kconrad/blurbs/analysis/entropypost.pdf.

In [BCM03, Cor. 1][9], 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

$$\mathcal{D} := \{b \in \mathbb{R}^m \mid F_b \cap \operatorname{dom} \Phi \neq \emptyset\} \tag{R12}$$
$$= \left\{ b \in \mathbb{R}^m \mid \exists x \in \operatorname{dom} \Phi, \ \forall i \in \overline{1, m} : \int_T x \psi_i d\mu = b_i \right\},$$

F_b being defined in (R6).

The situation is different in the general case.

[9] [BCM03] [J.M. Borwein, R. Choksi, P. Marechal: *Probability distributions of assets inferred from option prices via the principle of maximum entropy*, SIAM J. Optim. 14 (2003), 464–478.

An example

As an application of the previous approach, we consider

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

with $b = (b_1, \ldots, b_m) \in \mathbb{R}^m$,.

With our previous notation, $T := \mathbb{R}$, $\mathcal{A}$ is the class of Lebesgue measurable subsets of $\mathbb{R}$, μ is the Lebesgue measure, $\varphi = E_{BS}$ and $X := \mathcal{M}_0$; consequently, $\operatorname{dom}\Phi \subset \mathcal{M}_0^+$.

It follows that $0 \in \mathcal{D} \subset \mathcal{M}_0^+$, and so $b_{2k-1} \in \mathbb{R}_+$ if $b \in \mathcal{D}$ and $\mathbb{N}^* \ni k \leq (m+1)/2$, where $\mathcal{D}$ is defined in (R12).

The problem $(PG)_m$ is studied, for example, by Cover & Thomas [CT06] for $b_1 = 1$, and by Pavon & Ferrante [PF13] for $m = 3$ and $b = (1, 0, \sigma^2)$.

Proposition 7 (to be continued)

Consider the problem $(PG)_3$. Then

$$\mathcal{D} = \{(0,0,0)\} \cup \left\{b \in \mathbb{R}^3 \mid b_1 > 0,\ b_3 > 0,\ |b_2| < \sqrt{b_1 b_3}\right\}. \tag{R13}$$

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 $|b_2| < \sqrt{b_1 b_3}$ then the solution and the value of $(PG)_3$ are

$$\overline{x}(t) = \frac{b_1^2}{\sqrt{2\pi(b_1 b_3 - b_2^2)}} e^{-\frac{1}{2}\frac{b_1^2}{b_1 b_3 - b_2^2}(t - b_2)^2} \quad (t \in \mathbb{R}), \tag{R14}$$

$$\Phi(\overline{x}) = b_1 \ln \frac{b_1^2}{\sqrt{2\pi e(b_1 b_3 - b_2^2)}}, \tag{R15}$$

respectively.

Proposition 7 (continued)

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

$$\overline{x}(t) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{t^2}{2\sigma^2}} \quad (t \in \mathbb{R}), \qquad \Phi(\overline{x}) = -\ln \sqrt{2\pi e \sigma^2}. \qquad \text{(R16)}$$

Proof. Consider φ, X, Φ, ψ_k, X_k as above. For $x \in \mathcal{M}_0$, one has

$$\int_{\mathbb{R}} |tx(t)|\, dt \leq \sqrt{\int_{\mathbb{R}} |x(t)|\, dt} \cdot \sqrt{\int_{\mathbb{R}} t^2 |x(t)|\, dt}$$

by Hölder's inequality, with equality if and only if $x = 0$. Hence $\widetilde{X} = X_1 \cap X_3$; moreover, because $\mathcal{D} \subset \mathcal{M}_0^+$, from the above inequality we obtain that the inclusion $\subset$ holds in (R13).

Take $b_1, b_3 > 0$ and $|b_2| < \sqrt{b_1 b_3}$. The problem is to find (if possible) some $\overline{x} \in F_b \cap \operatorname{dom} \Phi$ ($\subset X \cap \widetilde{X}$) such that (R11) 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 all $t \in \mathbb{R}$. 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

$$\alpha = \frac{b_1^2}{b_1b_3 - b_2^2}, \quad \beta = \frac{b_2}{b_1}, \quad \gamma = \ln \frac{b_1^2}{\sqrt{2\pi(b_1b_3 - b_2^2)}}.$$

Hence $\overline{x}$ is the function defined in (R14). Consequently, $b := (b_1, b_2, b_3) \in \mathcal{D}$ and

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

Taking into account the constraints and the expressions of α, β, γ above, we get the formula for $\Phi(\overline{x})$ from (R15).

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 (R14) we get

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

which gives (R16) when $m = 0$.

Remark

The proof of the previous proposition shows that using LMM formally could be usefull not only for finding the optimal solutions of the entropy optimization problem (P), but also for finding the set S of those $b \in \mathbb{R}^m$ for which (P) has optimal solution.

As mentioned above, problem $(PG)_3$ is considered for $b = (1, 0, \sigma^2)$ by Pavon and Ferrante and solved applying [PF13, Cor. 9.3]. There $X = L_1(\mathbb{R})$, whence $X^* = L_\infty(\mathbb{R})$, and

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

It is not explained 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 7 provides an example in which $\mu(T) = \infty$ and the problem (P) has optimal solutions for all $b \in \mathcal{D}$. In [BCM03] 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 [VZ16] and [Z18] for $T := \mathbb{N}^*$, μ the counting measure, and $\varphi(u) = u \ln u - u$ for $u \geq 0$, $\varphi(u) = \infty$ for $u < 0$.

A complete study of (P) for $m = 1$ is given in [VZ16, Prop. 3.3]; so, besides providing the value of problem (P) for $b \in \mathcal{D}$ [$\mathcal{D}$ being defined in (R12)], when $\mathcal{D} \neq \{0\}$ it is shown that either (P) has optimal solution for each $b \in \mathcal{D}$, or
$(\operatorname{int} \mathcal{D}) \setminus \{b \in \mathcal{D} \mid (P) \text{ has optimal solution}\}$ is nonempty.

In [VZ16, 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 [Z, Th. 4.1]; the conclusions are similar to those in [VZ, Prop. 3.3] mentionted above.

Thank you for your attention!