ENTROPIC PROJECTIONS AND DOMINATING POINTS

Christian Léonard[1]

Abstract. Entropic projections and dominating points are solutions to convex minimization problems related to conditional laws of large numbers. They appear in many areas of applied mathematics such as statistical physics, information theory, mathematical statistics, ill-posed inverse problems or large deviation theory. By means of convex conjugate duality and functional analysis, criteria are derived for the existence of entropic projections, generalized entropic projections and dominating points. Representations of the generalized entropic projections are obtained. It is shown that they are the "measure component" of the solutions to some extended entropy minimization problem. This approach leads to new results and offers a unifying point of view. It also permits to extend previous results on the subject by removing unnecessary topological restrictions. As a by-product, new proofs of already known results are provided.

1. Introduction

Entropic projections and dominating points are solutions to convex minimization problems related to conditional laws of large numbers. They appear in many areas of applied mathematics such as statistical physics, information theory, mathematical statistics, ill-posed inverse problems or large deviation theory.

1.1. Conditional laws of large numbers

Suppose that the empirical measures

$$L_n := \frac{1}{n} \sum_{i=1}^{n} \delta_{Z_i}, \quad n \geq 1, \tag{1.1}$$

of the $\mathcal{Z}$-valued random variables $Z_1, Z_2, \ldots$ (δ_z is the Dirac measure at z) obey a Large Deviation Principle (LDP) in the set $P_{\mathcal{Z}}$ of all probability measures on $\mathcal{Z}$ with the rate function I. This approximately means that $\mathbb{P}(L_n \in \mathcal{A}) \underset{n \to \infty}{\asymp} \exp[-n \inf_{P \in \mathcal{A}} I(P)]$ for $\mathcal{A} \subset P_{\mathcal{Z}}$. With regular enough subsets $\mathcal{A}$ and $\mathcal{C}$ of $P_{\mathcal{Z}}$, we can expect that for "all" $\mathcal{A}$

$$\lim_{n \to \infty} \mathbb{P}(L_n \in \mathcal{A} \mid L_n \in \mathcal{C}) = \begin{cases} 1, & \text{if } \mathcal{A} \ni P_* \\ 0, & \text{otherwise} \end{cases} \tag{1.2}$$

where P_* is a minimizer of I on $\mathcal{C}$. To see this, remark that (formally) $\mathbb{P}(L_n \in \mathcal{A} \mid L_n \in \mathcal{C}) \underset{n \to \infty}{\asymp} \exp[-n(\inf_{P \in \mathcal{A} \cap \mathcal{C}} I(P) - \inf_{P \in \mathcal{C}} I(P))]$. A rigorous statement is given at Theorem 7.1.

If $(Z_1, \dots, Z_n)$ is exchangeable for each n, then (1.2) is equivalent to

$$\mathcal{L}(Z_1 \mid L_n \in \mathcal{C}) \underset{n \to \infty}{\longrightarrow} P_*$$

which means that conditionally on $L_n \in \mathcal{C}$, the law of any tagged "particle" (here we chose the first one) tends to P_* as n tends to infinity.

If I is strictly convex and $\mathcal{C}$ is convex, P_* is unique and (1.2) roughly means that conditionally on $L_n \in \mathcal{C}$, as n tends to infinity L_n tends to the solution P_* of the minimization problem

$$\text{minimize } I(P) \text{ subject to } P \in \mathcal{C}, \quad P \in P_{\mathcal{Z}}. \tag{1.3}$$

Such conditional Laws of Large Numbers (LLN) appear in information theory and in statistical physics where they are often called Gibbs conditioning principles (see [9], Sect. 7.3 and the references therein). If the variables Z_i are independent and identically distributed with law R, the LDP for the empirical measures is given by Sanov's theorem and the rate function I is the relative entropy

$$I(P|R) = \int_{\mathcal{Z}} \log\left(\frac{\mathrm{d}P}{\mathrm{d}R}\right) \mathrm{d}P, \quad P \in P_{\mathcal{Z}}.$$

Instead of the empirical probability measure of a random sample, one can consider another kind of random measure. Let $z_1, z_2, \dots$ be deterministic points in $\mathcal{Z}$ such that the empirical measure $\frac{1}{n} \sum_{i=1}^{n} \delta_{z_i}$ converges to $R \in P_{\mathcal{Z}}$. Let $W_1, W_2, \dots$ be a sequence of *independent* random real variables. The random measure of interest is

$$L_n = \frac{1}{n} \sum_{i=1}^{n} W_i \delta_{z_i} \tag{1.4}$$

where the W_i's are interpreted as random weights. If the weights are independent copies of W, as n tends to infinity, L_n tends to the deterministic measure $\mathbb{E}W.R$ and obeys the LDP in the space $M_{\mathcal{Z}}$ of measures on $\mathcal{Z}$ with rate function $I(Q) = \int_{\mathcal{Z}} \gamma^*(\frac{\mathrm{d}Q}{\mathrm{d}R}) \mathrm{d}R$, $Q \in M_{\mathcal{Z}}$ where γ^* is the Cramér transform of the law of W. In case the W_i's are not identically distributed, but have a law which depends (continuously) on z_i, one can again show that under additional assumptions L_n obeys the LDP in $M_{\mathcal{Z}}$ with rate function

$$I(Q) = \begin{cases} \int_{\mathcal{Z}} \gamma_z^*(\frac{\mathrm{d}Q}{\mathrm{d}R}(z)) \, R(\mathrm{d}z), & \text{if } Q \prec R \\ +\infty, & \text{otherwise,} \end{cases} \quad Q \in M_{\mathcal{Z}} \tag{1.5}$$

where γ_z^* is the Cramér transform of W_z. As γ^* is the convex conjugate of the log-Laplace transform of W, it is a convex function: I is a convex integral functional which is often called an *entropy*. Again, conditional LLNs hold for L_n and lead to the entropy minimization problem:

$$\text{minimize } I(Q) \text{ subject to } Q \in \mathcal{C}, \quad Q \in M_{\mathcal{Z}}. \tag{1.6}$$

The large deviations of these random measures and their conditional LLNs enter the framework of Maximum Entropy in the Mean (MEM) which has been studied among others by Csiszár et al. [7], Dacunha-Castelle and Gamboa [8], Gamboa and Gassiat [12] and Najim [21] and also [9] (Thm. 7.2.3). This problem also arises in the context of statistical physics. It has been studied among others by Boucher et al. [3] and Ellis et al. [11].

The relative entropy corresponds to $\gamma^*(t) = t \log t - t + 1$ in (1.5): the problem (1.3) is a special case of (1.6) with the additional constraint that $Q(\mathcal{Z}) = 1$.

In this paper, the constraint set $\mathcal{C}$ is assumed to be *convex* as is almost always done in the literature on the subject. This allows to rely on convex analysis, saddle-point theory or on the geometric theory of projection on convex sets.

The convex indicator of a subset A of X is denoted by

$$\iota_A(x) = \left\{ \begin{array}{ll} 0, & \text{if } x \in A \\ +\infty, & \text{otherwise,} \end{array} \right. \qquad x \in X. \tag{1.7}$$

Note that although ι_A is a convex function if and only if A is a convex set, we call it a *convex* indicator in any case to differentiate it from the probabilistic indicator $\mathbf{1}_A$.

1.2. Examples

We give a short list of popular entropies. They are described at (1.5) and characterized by their integrand γ^*.

(1) The relative entropy: $\gamma^*(t) = t \log t - t + 1 + \iota_{\{t \geq 0\}}$.
(2) The reverse relative entropy: $\gamma^*(t) = t - \log t - 1 + \iota_{\{t > 0\}}$.
(3) The Fermi-Dirac entropy: $\gamma^*(t) = \frac{1}{2}[(1 + t) \log(1 + t) + (1 - t) \log(1 - t)] + \iota_{\{-1 \leq t \leq 1\}}$.
(4) The L_p norm $(1 < p < \infty)$: $\gamma^*(t) = |t|^p / p$ and
(5) the L_p entropy $(1 < p < \infty)$: $\gamma^*(t) = t^p / p + \iota_{\{t \geq 0\}}$.

Note that the reverse relative entropy is $P \mapsto I(R|P)$.

In the case of the relative and reverse relative entropies, the global minimizer is R and one can interpret $I(P|R)$ and $I(R|P)$ as some type of distances between P and R. Note also that the positivity of the minimizers of (1.6) is guaranteed by $\{\gamma^* < \infty\} \subset [0, \infty)$. Consequently, the restriction that P is a probability measure is insured by the only unit mass constraint $P(\mathcal{Z}) = 1$.

A very interesting Bayesian interpretation of (1.6) is obtained in [8,12] in terms of the LDPs for L_n defined at (1.4). The above entropies correspond to iid weights W distributed as follows:

(1) the relative entropy: $\text{Law}(W) = \text{Poisson}(1)$,
(2) the reverse relative entropy: $\text{Law}(W) = \text{Exponential}(1)$,
(3) the Fermi-Dirac entropy: $\text{Law}(W) = (\delta_{-1} + \delta_{+1})/2$,
(4) the L_2 norm: $\text{Law}(W) = \text{Normal}(0,1)$.

1.3. Entropic and generalized entropic projections

The minimizers of (1.6) are called *entropic projections*. It may happen that even if the minimizer is not attained, any minimizing sequence converges to some measure Q_* which does not belong to $\mathcal{C}$. This intriguing phenomenon was discovered by Csiszár [5]. Such a Q_* is called a *generalized* entropic projection.

In the special case where I is the relative entropy, Csiszár has obtained existence results in [4] together with dual equalities. His proofs are based on geometric properties of the relative entropy; no convex analysis is needed. Based on the same geometric ideas, he obtained later in [5] a powerful Gibbs conditioning principle for noninteracting particles. For general entropies as in (1.5), he studies the problem of existence of entropic and generalized entropic projections in [6].

The minimization problem (1.6) is interesting in its own right, even when conditional LLNs are not at stake. The literature on this subject is huge, for instance see [13,29] and the references therein.

1.4. The extended minimization problem

As will be seen, it is worth introducing an *extension* of (1.6) to take the *generalized* projections into account. A solution to (1.6) is in $A_{\mathcal{Z}}$: the space of all $Q \in M_{\mathcal{Z}}$ which are absolutely continuous with respect to R. We consider an extension $\bar{I}$ of the entropy I to a vector space $L_{\mathcal{Z}}$ which is the direct sum $L_{\mathcal{Z}} = A_{\mathcal{Z}} \oplus S_{\mathcal{Z}}$ of $A_{\mathcal{Z}}$ and a vector space $S_{\mathcal{Z}}$ of *singular linear forms* (acting on numerical functions) which may not be σ-additive.

Any ℓ in $L_{\mathcal{Z}}$ is uniquely decomposed into $\ell = \ell^a + \ell^s$ with $\ell^a \in A_{\mathcal{Z}}$ and $\ell^s \in S_{\mathcal{Z}}$ and $\bar{I}$ has the following shape

$$\bar{I}(\ell) = I(\ell^a) + I^s(\ell^s)$$

where I^s is a positively homogeneous function on $S_{\mathcal{Z}}$. See (2.9) for the precise description of $\bar{I}$. For instance, the extended relative entropy is

$$\bar{I}(\ell|R) = I(\ell^a|R) + \sup\left\{ \langle \ell^s, u\rangle; u, \int_{\mathcal{Z}} e^u \, dR < \infty \right\}, \quad \ell \in L_{\mathcal{Z}}$$

and actually $\langle \ell^s, u\rangle = 0$ for any ℓ such that $\bar{I}(\ell) < \infty$ and any u such that $\int_{\mathcal{Z}} e^{a|u|} \, dR < \infty$ for *all* $a > 0$. The reverse entropy, L_1-norm and L_1-entropy also admit nontrivial extensions. On the other hand, the extensions of the Fermi-Dirac, L_p-norm and L_p-entropy with $p > 1$ are trivial: $\{k \in S_{\mathcal{Z}}; I^s(k) < \infty\} = \{0\}$.

The extended problem is

$$\text{minimize } \bar{I}(\ell) \text{ subject to } \ell \in \mathcal{C}, \quad \ell \in L_{\mathcal{Z}}. \tag{1.8}$$

In fact, $\bar{I}$ is chosen to be the largest convex lower semicontinuous extension of I to $L_{\mathcal{Z}}$ with respect to some weak topology. This guarantees tight relations between (1.6) and (1.8). In particular, one can expect that their values are equal for a large class of convex sets $\mathcal{C}$.

Even if I is strictly convex, $\bar{I}$ isn't strictly convex in general since I^s is positively homogeneous, so that (1.8) may admit several minimizers.

Examples will be given of interesting situations where (1.6) is not attained in $A_{\mathcal{Z}}$ while (1.8) is attained in $L_{\mathcal{Z}}$.

1.5. Dominating points

Let the constraint set $\mathcal{C}$ be described by

$$\mathcal{C} = \{Q \in M_{\mathcal{Z}}; TQ \in C\} \tag{1.9}$$

where C is a subset of a vector space $\mathcal{X}$ and $T : M_{\mathcal{Z}} \to \mathcal{X}$ is a linear operator. As a typical example, one can think of

$$TQ = \int_{\mathcal{Z}} \theta(z) \, Q(dz) \tag{1.10}$$

where $\theta : \mathcal{Z} \to \mathcal{X}$ is some function and the integral should be taken formally for the moment. With L_n given at (1.1) or (1.4), if T is regular enough, we obtain by the contraction principle that $X_n := TL_n = \frac{1}{n}\sum_{i=1}^n \theta(Z_i) \in \mathcal{X}$ or $X_n := TL_n = \frac{1}{n}\sum_{i=1}^n W_i\theta(z_i)$ obeys the LDP in $\mathcal{X}$ with rate function $J(x) = \inf\{I(Q); Q \in M_{\mathcal{Z}}, TQ = x\}$, $x \in \mathcal{X}$. Once again, the conditional LLN for X_n is of the form: For "all" $A \subset \mathcal{X}$,

$$\lim_{n\to\infty} \mathbb{P}(X_n \in A \mid X_n \in C) = \begin{cases} 1, & \text{if } A \ni x_* \\ 0, & \text{otherwise} \end{cases}$$

where x_* is a solution to the minimization problem

$$\text{minimize } J(x) \text{ subject to } x \in C, \quad x \in \mathcal{X}. \tag{1.11}$$

The minimizers of (1.11) are called *dominating points.* This notion was introduced by Ney [22,23] in the special case where $(Z_i)_{i\geq 1}$ is an iid sequence in $\mathcal{Z} = \mathbb{R}^d$ and θ is the identity, *i.e.* $X_n = \frac{1}{n}\sum_{i=1}^n Z_i$. Later, Einmahl and Kuelbs [10,14] have extended this study to a Banach space $\mathcal{Z}$. In this iid case, J is the Cramér transform of the law of Z_1.

1.6. **Presentation of the results**

We treat the problems of existence of entropic projections and dominating points in a unified way, taking advantage of the mapping $TQ = x$. Hence, we mainly concentrate on the entropic projections and then transport the results to the dominating points.

It will be proved at Proposition 5.4 that the entropic projection exists on $\mathcal{C}$ if the supporting hyperplanes of $\mathcal{C}$ are directed by sufficiently integrable functions. In some cases of not enough integrable supporting hyperplanes, the representation of the generalized projection is still available and given at Theorem 5.6.

It will appear that the generalized projection is the "measure" part of the minimizers of (1.8). For instance, with the relative entropy $I(.|R)$, the projection exists in $\mathcal{C}$ if its supporting hyperplanes are directed by functions u such that $\int_{\mathcal{Z}} e^{\alpha|u|} \, dR < \infty$ for *all* $\alpha > 0$, see Proposition 5.7. If these u only satisfy $\int_{\mathcal{Z}} e^{\alpha|u|} \, dR < \infty$ for *some* $\alpha > 0$, the projection may not exist in $\mathcal{C}$, but the generalized projection is computable: its Radon-Nykodym derivative with respect to R is characterized at Proposition 5.8.

We find again some already known results of Csiszár [5,6], Einmahl and Kuelbs [10,14] with different proofs and a new point of view. The representations of the generalized projections are new results. The conditions on C to obtain dominating points are improved and an interesting phenomenon noticed in [14] is clarified at Remark 6.8 by connecting it with the generalized entropic projection.

Finally, a probabilistic interpretation of the singular components of the generalized projections is proposed at Section 7. It is obtained in terms of conditional LLNs.

The main results are Theorems 4.1, 4.4, 5.6 and 6.7.

1.7. **Outline of the paper**

At Section 2, we give precise formulations of the entropy minimization problems (1.6) and (1.8). Then we recall at Theorems 2.7 and 2.8 results from [19] about the existence and uniqueness of the solutions of these problems, related dual equalities and the characterizations of their solutions in terms of integral representations.

Examples of standard entropies and constraints are presented at Section 3.

We show at Theorem 4.4 in Section 4 that under "critical" constraints, although the problem (1.6) may not be attained, its minimizing sequences may converge in some sense to some measure Q_*: the *generalized* entropic projection.

Section 5 is mainly a restatement of Sections 2 and 4 in terms of entropic projections. The results are also stated explicitly for the special important case of the relative entropy.

Section 6 is devoted to dominating points. As they are continuous images of entropic projections, the main results of this section are corollaries of the results of Section 5.

At Section 7, we illustrate our results in terms of conditional LLNs, see (1.2). In particular, the generalized projections and singular components of the minimizers of (1.8) are interpreted in terms of these conditional LLNs.

1.8. **Notation**

Let X and Y be topological vector spaces. The algebraic dual space of X is X^*, the topological dual space of X is X'. The topology of X weakened by Y is $\sigma(X, Y)$ and we write $\langle X, Y \rangle$ to specify that X and Y are in separating duality.

Let $f : X \to [-\infty, +\infty]$ be an extended numerical function. Its convex conjugate with respect to $\langle X, Y \rangle$ is

$$f^*(y) = \sup_{x \in X}\{\langle x, y \rangle - f(x)\} \in [-\infty, +\infty], \quad y \in Y. \tag{1.12}$$

Its subdifferential at x with respect to $\langle X, Y \rangle$ is

$$\partial_Y f(x) = \{y \in Y; f(x + \xi) \geq f(x) + \langle y, \xi \rangle, \forall \xi \in X\}.$$

If no confusion occurs, we write $\partial f(x)$.

Let A be a subset of X, its intrinsic core is icor $A = \{x \in A; \forall x' \in \mathrm{aff}\, A, \exists t > 0, [x, x + t(x' - x)] \subset A\}$ where $\mathrm{aff}\, A$ is the affine space spanned by A. Let us denote $\mathrm{dom}\, f = \{x \in X; f(x) < \infty\}$ the effective domain of f and $\mathrm{icordom}\, f$ the intrinsic core of $\mathrm{dom}\, f$.

The convex indicator ι_A of a subset A of X is defined at (1.7).

We write

$$I_\varphi(u) := \int_{\mathcal{Z}} \varphi(z, u(z))\, R(\mathrm{d}z) = \int_{\mathcal{Z}} \varphi(u)\, \mathrm{d}R$$

and $I = I_{\gamma^*}$ for short, instead of (1.5).

The inf-convolution of f and g is $f \square g(z) = \inf\{f(x) + g(y); x, y : x + y = z\}$.

An index of notation is provided at the end of the article.

2. Minimizing entropy under convex constraints

In this section, the main results of [19] are recalled. They are Theorems 2.7 and 2.8 below and their statements necessitate the notion of Orlicz spaces. First, we recall basic definitions and notions about these function spaces which are natural extensions of the standard L_p spaces.

2.1. **Orlicz spaces**

The fact that the generalized projection may not belong to $\mathcal{C}$ is connected with some properties of Orlicz spaces associated to I. Let us recall some basic definitions and results. A standard reference is [24].

A set $\mathcal{Z}$ is furnished with a σ-finite nonnegative measure R on a σ-field which is assumed to be R-complete. A function $\rho : \mathcal{Z} \times \mathbb{R}$ is said to be a *Young function* if for R-almost every z, $\rho(z, \cdot)$ is a convex even $[0, \infty]$-valued function on $\mathbb{R}$ such that $\rho(z, 0) = 0$ and there exists a measurable function $z \mapsto s_z > 0$ such that $0 < \rho(z, s_z) < \infty$.

In the sequel, every numerical function on $\mathcal{Z}$ is supposed to be measurable.

Definition 2.1 (the Orlicz spaces L_ρ and E_ρ). The *Orlicz space* associated with ρ is defined by $L_\rho = \{u : \mathcal{Z} \to \mathbb{R}; \|u\|_\rho < +\infty\}$ where the Luxemburg norm $\|\cdot\|_\rho$ is defined by

$$\|u\|_\rho = \inf \left\{ \beta > 0\;;\; \int_{\mathcal{Z}} \rho(z, u(z)/\beta)\, R(\mathrm{d}z) \leq 1 \right\}$$

and R-a.e. equal functions are identified. Hence,

$$L_\rho = \left\{ u : \mathcal{Z} \to \mathbb{R}\;; \exists \alpha_o > 0, \int_{\mathcal{Z}} \rho\Big(z, \alpha_o u(z)\Big)\, R(\mathrm{d}z) < \infty \right\}.$$

A subspace of interest is

$$E_\rho = \left\{ u : \mathcal{Z} \to \mathbb{R}\;; \forall \alpha > 0, \int_{\mathcal{Z}} \rho\Big(z, \alpha u(z)\Big)\, R(\mathrm{d}z) < \infty \right\}.$$

Taking $\rho_p(z, s) = |s|^p/p$ with $p \geq 1$ gives $L_{\rho_p} = E_{\rho_p} = L_p$ and the corresponding Luxemburg norm is $\|\cdot\|_{\rho_p} = p^{-1/p}\|\cdot\|_p$ where $\|\cdot\|_p$ is the usual L_p norm.

The convex conjugate ρ^* of a Young function is still a Young function, so that L_{ρ^*} is also an Orlicz space. Hölder's inequality in Orlicz spaces is

$$\|uv\|_1 \leq 2\|u\|_\rho \|v\|_{\rho^*}, \quad u \in L_\rho, v \in L_{\rho^*}. \tag{2.1}$$

For instance, with $\rho_p^*(t) = |t|^q/q$, $1/p + 1/q = 1$, (2.1) is $\|uv\|_1 \leq 2p^{-1/p} q^{-1/q}\|u\|_p \|v\|_q$. Note that $2p^{-1/p} q^{-1/q} \geq 1$ with equality when $p = 2$.

Theorem 2.2 (representation of E'_ρ). *Suppose that ρ is a finite Young function. Then, the dual space of E_ρ is isomorphic to L_{ρ^*}, the Orlicz space associated with the Young function ρ^* which is the convex conjugate of ρ.*

A continuous linear form $\ell \in L'_\rho$ is said to be *singular* if for all $u \in L_\rho$, there exists a decreasing sequence of measurable sets (A_n) such that $R(\cap_n A_n) = 0$ and for all $n \geq 1$, $\langle \ell, u\mathbf{1}_{\mathcal{Z}\setminus A_n}\rangle = 0$.

Proposition 2.3. *Let us assume that ρ is finite. Then, $\ell \in L'_\rho$ is singular if and only if $\langle \ell, u\rangle = 0$, for all u in E_ρ.*

Let us denote respectively $L_{\rho^*}R = \{fR; f \in L_{\rho^*}\}$ and L^s_ρ the subspaces of L'_ρ of all absolutely continuous and singular forms.

Theorem 2.4 (representation of L'_ρ). *Let ρ be any Young function. The dual space of L_ρ is isomorphic to the direct sum $L'_\rho = L_{\rho^*}R \oplus L^s_\rho$. This implies that any $\ell \in L'_\rho$ is uniquely decomposed as*

$$\ell = \ell^a + \ell^s \tag{2.2}$$

with $\ell^a \in L_{\rho^*}R$ and $\ell^s \in L^s_\rho$.

In the decomposition (2.2), ℓ^a is called the *absolutely continuous* part of ℓ while ℓ^s is its *singular part*.
The function ρ is said to satisfy the Δ_2-condition if

$$\text{there exist } \kappa > 0, s_o \geq 0 \text{ such that } \forall s \geq s_o, z \in \mathcal{Z}, \rho_z(2s) \leq \kappa\rho_z(s). \tag{2.3}$$

When R is bounded, in order that $E_\rho = L_\rho$, it is enough that ρ satisfies the Δ_2-condition. Consequently, in this situation we have $L'_\rho = L_{\rho^*}R$ so that L^s_ρ reduces to the null vector space.

We shall see below that the important case of the relative entropy leads to a Young function ρ which doesn't satisfy the Δ_2-condition; non-trivial singular components will appear.

2.2. **The assumptions**

Let R be a positive measure on a space $\mathcal{Z}$ and take a $[0,\infty]$-valued measurable function γ^* on $\mathcal{Z}\times\mathbb{R}$ to define I as at (1.5). In order to define the constraint in a way similar to (1.10), take a function

$$\theta : \mathcal{Z} \to \mathcal{X}_o$$

where $\mathcal{X}_o$ is the algebraic dual space of some vector space $\mathcal{Y}_o$. Let us collect the assumptions on R, γ^*, θ and C.

2.3. **Assumptions (A).**

(A_R) It is assumed that the reference measure R is a bounded positive measure on a space $\mathcal{Z}$ endowed with some R-complete σ-field.
(A_C) C is a convex subset of $\mathcal{X}_o$.
(A_{γ^*}) *Assumptions on γ^*.*
    (1) $\gamma^*(\cdot, t)$ is z-measurable for all t and for R-almost every $z \in \mathcal{Z}$, $\gamma^*(z, \cdot)$ is a lower semicontinuous strictly convex $[0,+\infty]$-valued function on $\mathbb{R}$.
    (2) It is also assumed that for R-almost every $z \in \mathcal{Z}$, $\gamma^*(z, \cdot)$ attains a unique minimum denoted by $m(z)$, the minimum value is $\gamma^*_z(m(z)) = 0$, $\forall z \in \mathcal{Z}$ and there exist $a(z), b(z) > 0$ such that $0 < \gamma^*(z, m(z) + a(z)) < \infty$ and $0 < \gamma^*(z, m(z) - b(z)) < \infty$.
    (3) $\int_{\mathcal{Z}} \gamma^*((1+\alpha)m)\,dR + \int_{\mathcal{Z}} \gamma^*((1-\alpha)m)\,dR < \infty$, for some $\alpha > 0$.
The function γ^*_z is the convex conjugate of the lower semicontinuous convex function $\gamma_z = \gamma^{**}_z$. Defining

$$\lambda(z, s) = \gamma(z, s) - m(z)s, \quad z \in \mathcal{Z}, s \in \mathbb{R}, \tag{2.4}$$

we see that for R-a.e. z, λ_z is a nonnegative convex function and it vanishes at $s = 0$.

(A_θ) *Assumptions on θ.*
  (1) for any $y \in \mathcal{Y}_o$, the function $z \in \mathcal{Z} \mapsto \langle y, \theta(z) \rangle \in \mathbb{R}$ is measurable;
  (2) for any $y \in \mathcal{Y}_o$, $\langle y, \theta(\cdot) \rangle = 0, R$-a.e. implies that $y = 0$;
  ($\exists$) $\forall y \in \mathcal{Y}_o, \exists \alpha > 0, \quad \int_{\mathcal{Z}} \lambda(\alpha \langle y, \theta \rangle) \, \mathrm{d}R < \infty.$

Since

$$\lambda_\diamond(z, s) = \max[\lambda(z, s), \lambda(z, -s)] \in [0, \infty], \quad z \in \mathcal{Z}, s \in \mathbb{R} \tag{2.5}$$

is a Young function, one can consider the corresponding Orlicz spaces $L_{\lambda_\diamond}$ and $L_{\lambda_\diamond^*}$ where $\lambda_\diamond^*(z, \cdot)$ is the convex conjugate of $\lambda_\diamond(z, \cdot)$.

Remark 2.5 (some comments about these assumptions)**.**
  (1) Assuming R to be σ-finite is standard when working with integral functionals. Here, R is assumed to be bounded for simplicity.
  (2) Thanks to the theory of normal integrands [25], $\gamma = \gamma^{**}$ is jointly measurable.
    The function m is measurable since its graph $\{(z, u) : u = m(z)\}$ is measurable. Indeed, it is the (z, u)-projection of the intersection of the horizontal hyperplane $u = 0$ with the measurable graph $\{(t, z, u) : \gamma_z^*(t) = u\}$ of γ^*. Note that we use the uniqueness assumption in ($A_{\gamma^*}^2$) to avoid a measurable selection argument.
    As a consequence, λ is a jointly measurable function.
  (3) The assumption ($A_{\gamma^*}^3$) implies that m is in the Orlicz space $L_{\lambda_\diamond^*}$, see (2.5). In turn, the centered form $\ell - mR$ is in $L_{\lambda_\diamond}'$ whenever ℓ is. This will be used later without warning. It also follows from (2.4) that $\gamma_z^*(t) = \lambda_z^*(t - m(z))$. These two facts allow us to write

$$I_{\gamma^*}(fR) = I_{\lambda^*}(fR - mR), \quad f \in L_{\lambda_\diamond^*}$$

    where λ_z^* is the convex conjugate of λ_z.
  (4) The assumption ($A_\theta^\exists$) is equivalent to $\langle y, \theta(\cdot) \rangle \in L_{\lambda_\diamond}$ for all $y \in \mathcal{Y}_o$, see (2.7) below.

2.4. **Examples**

We consider the entropies which were introduced after (1.7) and give the corresponding functions λ and $\lambda_\diamond$.
  (1) The relative entropy: $\lambda(s) = \mathrm{e}^s - s - 1, \lambda_\diamond(s) = \lambda(|s|)$.
  (2) The reverse relative entropy: $\lambda(s) = -\log(1 - s) - s + \iota_{\{s < 1\}}, \lambda_\diamond(s) = \lambda(|s|)$.
  (3) The Fermi-Dirac entropy: $\lambda(s) = \lambda_\diamond(s) = \log \cosh s$.
  (4) The L_p norm ($1 < p < \infty$): $\lambda(s) = \lambda_\diamond(s) = |s|^q / q$ with $1/p + 1/q = 1$ and
  (5) the L_p entropy ($1 < p < \infty$): $\lambda(s) = \begin{cases} 0 & \text{if } s \leq 0 \\ s^q/q & \text{if } s \geq 0, \end{cases} \quad \lambda_\diamond(s) = |s|^q/q$.

2.5. **The entropy minimization problems ($\mathbf{P}_C$) and ($\overline{\mathbf{P}}_C$)**

Because of the Remark 2.5(3), under ($A_{\gamma^*}^3$) the effective domain of I given at (1.5) is included in $L_{\lambda_\diamond^*}R$ and the entropy functional to be considered is defined by

$$I(fR) = \int_{\mathcal{Z}} \gamma^*(f) \, \mathrm{d}R, \quad f \in L_{\lambda_\diamond^*}. \tag{2.6}$$

Assuming (A_θ), that is

$$\langle \mathcal{Y}_o, \theta(\cdot) \rangle \subset L_{\lambda_\diamond}, \tag{2.7}$$

Hölder's inequality in Orlicz spaces, see (2.1), allows to define the constraint operator $T : \ell \in L_{\lambda_\diamond}' \mapsto \langle \theta, \ell \rangle \in \mathcal{X}_o$ by:

$$\left\langle y, \langle \theta, \ell \rangle \right\rangle_{\mathcal{Y}_o, \mathcal{X}_o} = \left\langle \langle y, \theta \rangle, \ell \right\rangle_{L_{\lambda_\diamond}, L_{\lambda_\diamond}'}, \forall y \in \mathcal{Y}_o. \tag{2.8}$$

If $\ell = Q \in L_{\lambda_\diamond^*}R \subset M_{\mathcal{Z}}$, one writes $TQ = \langle \theta, Q \rangle = \int_{\mathcal{Z}} \theta \, \mathrm{d}Q$ to mean (2.8); this is a $\mathcal{X}_o$-valued weak integral.

Example 2.6 (moment constraint)**.** This is the easiest constraint to think of. Let $\theta = (\theta_k)_{1\le k\le K}$ be a measurable function from $\mathcal{Z}$ to $\mathcal{X}_o = \mathbb{R}^K$. The moment constraint is specified by the operator

$$\int_{\mathcal{Z}} \theta \, \mathrm{d}\ell = \left(\int_{\mathcal{Z}} \theta_k \, \mathrm{d}\ell \right)_{1\le k\le K} \in \mathbb{R}^K,$$

which is defined for each $\ell \in M_{\mathcal{Z}}$ which integrates all the real valued measurable functions θ_k.

The minimization problem (1.6) becomes

$$\textit{minimize } I(Q) \textit{ subject to } \int_{\mathcal{Z}} \theta \, \mathrm{d}Q \in C, \quad Q \in L_{\lambda_\diamond^*} R \tag{P_C}$$

where C is a convex subset of $\mathcal{X}_o$. The extended entropy is defined by

$$\bar{I}(\ell) = I(\ell^a) + I^s(\ell^s), \quad \ell \in L'_{\lambda_\diamond} \tag{2.9}$$

where, using the notation of Theorem 2.4,

$$I^s(\ell^s) = \iota^*_{\mathrm{dom}\, I_\gamma}(\ell^s) = \sup\left\{ \langle \ell^s, u \rangle; u \in L_{\lambda_\diamond}, I_\gamma(u) < \infty \right\} \in [0, \infty]. \tag{2.10}$$

It is proved in [18] that $\bar{I}$ is the greatest convex $\sigma(L'_{\lambda_\diamond}, L_{\lambda_\diamond})$-lower semicontinuous extension of I to $L'_{\lambda_\diamond} \supset L_{\lambda_\diamond^*} R$. The associated extended minimization problem is

$$\textit{minimize } \bar{I}(\ell) \textit{ subject to } \langle \theta, \ell \rangle \in C, \quad \ell \in L'_{\lambda_\diamond}. \tag{$\overline{\mathrm{P}}_C$}$$

2.6. **Good and critical constraints**

If the Young function $\lambda_\diamond$ doesn't satisfy the Δ_2-condition (2.3) as for instance with the relative entropy and the reverse relative entropy (see items (1) and (2) of the examples above), the *small* Orlicz space $E_{\lambda_\diamond}$ may be a proper subset of $L_{\lambda_\diamond}$. Consequently, for some functions θ, the integrability property

$$\langle \mathcal{Y}_o, \theta(\cdot) \rangle \subset E_{\lambda_\diamond} \tag{2.11}$$

or equivalently

$$\forall y \in \mathcal{Y}_o, \int_{\mathcal{Z}} \lambda(\langle y, \theta \rangle) \, \mathrm{d}R < \infty \tag{$\mathrm{A}_\theta^\forall$}$$

may not be satisfied while the weaker property

$$\forall y \in \mathcal{Y}_o, \exists \alpha > 0, \int_{\mathcal{Z}} \lambda(\alpha \langle y, \theta \rangle) \, \mathrm{d}R < \infty \tag{$\mathrm{A}_\theta^\exists$}$$

which is equivalent to (2.7), holds. In this situation, analytical complications occur, see Section 4. This is the reason why constraint satisfying $(\mathrm{A}_\theta^\forall)$ are called *good constraints*, while constraints satisfying $(\mathrm{A}_\theta^\exists)$ but not $(\mathrm{A}_\theta^\forall)$ are called *critical constraints*.

2.7. **Definitions of $\mathcal{Y}, \mathcal{X}, T^*, \Gamma^*$ and (D_C)**

These objects will be necessary to state the relevant dual problems. The general hypotheses (A) are assumed.

The space $\mathcal{Y}$. Because of the hypotheses (A_θ^2) and $(\mathrm{A}_\theta^\exists)$, $\mathcal{Y}_o$ can be identified with the subspace $\langle \mathcal{Y}_o, \theta(\cdot) \rangle$ of $L_{\lambda_\diamond}$. The space $\mathcal{Y}$ is the extension of $\mathcal{Y}_o$ which is isomorphic to the $\| \cdot \|_{\lambda_\diamond}$-closure of $\langle \mathcal{Y}_o, \theta(\cdot) \rangle$ in $L_{\lambda_\diamond}$.

The space $\mathcal{X}$. The topological dual space of $\mathcal{Y}$ is $\mathcal{X} = \mathcal{Y}' \subset \mathcal{X}_o$. $\mathcal{X}$ is identified with $L'_{\lambda_\diamond}/\ker T$.

Since $\lambda_\diamond$ is finite, we can apply Proposition 2.3. It tells us that under the assumption $(A^\forall_\theta)$, $TL^s_{\lambda_\diamond} = \{0\}$ so that $\mathcal{X} \cong L_{\lambda^*_\diamond}R/\ker T$.

The operator T^*. Let us define the adjoint $T^\sharp : \mathcal{X}^* \to L'^*_{\lambda_\diamond}$ for all $\omega \in \mathcal{X}^*$ by: $\langle \ell, T^\sharp \omega \rangle_{L'_{\lambda_\diamond}, L'^*_{\lambda_\diamond}} = \langle T\ell, \omega \rangle_{\mathcal{X}, \mathcal{X}^*}, \forall \ell \in L'_{\lambda_\diamond}$. We have the inclusions $\mathcal{Y}_o \subset \mathcal{Y} \subset \mathcal{X}^*$. The adjoint operator T^* is the restriction of $T^\sharp$ to $\mathcal{Y}_o$. With some abuse of notation, we still denote $T^*y = \langle y, \theta \rangle$ for $y \in \mathcal{Y}$. This can be interpreted as a dual bracket between $\mathcal{X}^*_o$ and $\mathcal{X}_o$ since $T^*y = \langle \tilde{y}, \theta \rangle$ R-a.e. for some $\tilde{y} \in \mathcal{X}^*_o$.

The function Γ^*. The basic dual problem associated with (P_C) and $(\overline{\mathrm{P}}_C)$ is

$$maximize \ \inf_{x \in C} \langle y, x \rangle - \Gamma(y), \quad y \in \mathcal{Y}_o$$

where

$$\Gamma(y) = I_\gamma(\langle y, \theta \rangle), \quad y \in \mathcal{Y}_o. \tag{2.12}$$

Let us denote

$$\Gamma^*(x) = \sup_{y \in \mathcal{Y}_o} \{\langle y, x \rangle - I_\gamma(\langle y, \theta \rangle)\}, \quad x \in \mathcal{X}_o \tag{2.13}$$

its convex conjugate. It is shown in [16] (Sect. 4), that $\mathrm{dom}\,\Gamma^* \subset \mathcal{X}$.

The dual problem (D_C). Another dual problem associated with (P_C) and $(\overline{\mathrm{P}}_C)$ is

$$maximize \ \inf_{x \in C \cap \mathcal{X}} \langle y, x \rangle - I_\gamma(\langle y, \theta \rangle), \quad y \in \mathcal{Y}. \tag{D_C}$$

2.8. Solving (P_C)

In this section, we study (P_C) under the good constraint hypothesis $(A^\forall_\theta)$ which imposes that $T^*\mathcal{Y} \subset E_{\lambda_\diamond}$ and $T(L_{\lambda^*_\diamond}R) \subset \mathcal{X}$.

The extended dual problem $(\widetilde{\mathrm{D}}_C)$. The extended dual problem associated with (P_C) is

$$maximize \ \inf_{x \in C \cap \mathcal{X}} \langle \omega, x \rangle - I_\gamma(\langle \omega, \theta \rangle), \quad \omega \in \widetilde{\mathcal{Y}} \tag{$\widetilde{\mathrm{D}}_C$}$$

where $\widetilde{\mathcal{Y}} \subset \mathcal{X}^*$ is some cone which contains $\mathcal{Y}_o$. Its exact definition is given at Appendix A.

Theorem 2.7. *Suppose that*

    (1) the hypotheses (A) *and* $(A^\forall_\theta)$ *are satisfied;*
    (2) the convex set C is assumed to be such that

$$T^{-1}C \cap L_{\lambda^*_\diamond}R = \bigcap_{y \in Y} \left\{ fR \in L_{\lambda^*_\diamond}R; \int_{\mathcal{Z}} \langle y, \theta \rangle f \, dR \geq a_y \right\} \tag{2.14}$$

    for some subset $Y \in \mathcal{X}^*_o$ with $\langle y, \theta \rangle \in E_{\lambda_\diamond}$ for all $y \in Y$ and some function $y \in Y \mapsto a_y \in \mathbb{R}$. In other words, $T^{-1}C \cap L_{\lambda^*_\diamond}R$ is a $\sigma(L_{\lambda^*_\diamond}R, E_{\lambda_\diamond})$-closed convex subset of $L_{\lambda^*_\diamond}R$.

Then:

(a) The dual equality for (P_C) is

$$\inf(\mathrm{P}_C) = \sup(\mathrm{D}_C) = \sup(\widetilde{\mathrm{D}}_C) = \inf_{x \in C} \Gamma^*(x) \in [0, \infty].$$

(b) If $C \cap \mathrm{dom}\,\Gamma^* \neq \emptyset$ or equivalently $C \cap T\mathrm{dom}\,I \neq \emptyset$, then (P_C) admits a unique solution $\widehat{Q}$ in $L_{\lambda^*_\diamond}R$ and any minimizing sequence $(Q_n)_{n \geq 1}$ converges to $\widehat{Q}$ with respect to the topology $\sigma(L_{\lambda^*_\diamond}R, L_{\lambda_\diamond})$.
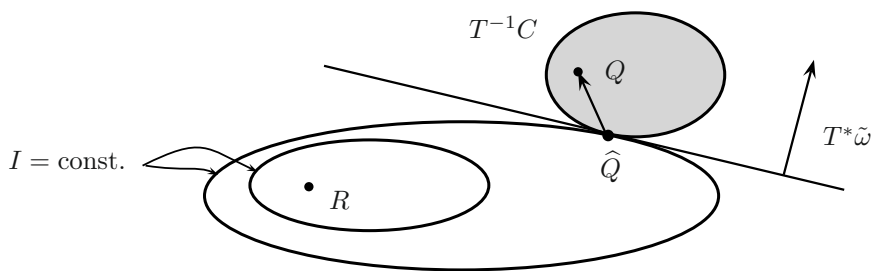
FIGURE 1.

Suppose that in addition $C \cap \operatorname{icordom} \Gamma^* \neq \emptyset$ or equivalently $C \cap \operatorname{icor}(T \operatorname{dom} I) \neq \emptyset$.

(c) Let us define $\hat{x} := \int_{\mathcal{Z}} \theta \, \mathrm{d}\widehat{Q}$ in the weak sense with respect to the duality $\langle \mathcal{Y}_o, \mathcal{X}_o \rangle$. There exists $\tilde{\omega} \in \widetilde{\mathcal{Y}}$ such that

$$\begin{cases} (a) & \hat{x} \in C \cap \operatorname{dom} \Gamma^* \\ (b) & \langle \tilde{\omega}, \hat{x} \rangle_{\mathcal{X}^*, \mathcal{X}} \leq \langle \tilde{\omega}, x \rangle_{\mathcal{X}^*, \mathcal{X}}, \forall x \in C \cap \operatorname{dom} \Gamma^* \\ (c) & \widehat{Q}(\mathrm{d}z) = \gamma_z'(\langle \tilde{\omega}, \theta(z) \rangle) \, R(\mathrm{d}z). \end{cases} \tag{2.15}$$

Furthermore, $\widehat{Q} \in L_{\lambda_\diamond^*} R$ and $\tilde{\omega} \in \widetilde{\mathcal{Y}}$ satisfy (2.15) if and only if $\widehat{Q}$ solves (P_C) and $\tilde{\omega}$ solves $(\widetilde{\mathrm{D}}_C)$.

(d) Of course, $(2.15)c$ implies

$$\hat{x} = \int_{\mathcal{Z}} \theta \gamma'(\langle \tilde{\omega}, \theta \rangle) \, \mathrm{d}R \tag{2.16}$$

in the weak sense. Moreover,
1. $\hat{x}$ minimizes Γ^* on C;
2. $I(\widehat{Q}) = \Gamma^*(\hat{x}) = \int_{\mathcal{Z}} \gamma^* \circ \gamma'(\langle \tilde{\omega}, \theta \rangle) \, \mathrm{d}R < \infty$; and
3. $I(\widehat{Q}) + \int_{\mathcal{Z}} \gamma(\langle \tilde{\omega}, \theta \rangle) \, \mathrm{d}R = \int_{\mathcal{Z}} \langle \tilde{\omega}, \theta \rangle \, \mathrm{d}\widehat{Q}$.

Proof. This result is [19] (Thm. 3.2). □

Figure 1 illustrates the items (a) and (b) of (2.15) which, with $T\widehat{Q} = \hat{x}, TQ = x \in C$, can be rewritten: $\widehat{Q} \in T^{-1}C$ and $\langle T^*\tilde{\omega}, Q - \widehat{Q} \rangle \geq 0, \forall Q \in T^{-1}C$: the shaded area.

As R is the unconstrained minimizer of I, the level sets of I form an increasing family of convex sets which expends from R. One sees that the hyperplane $\{Q : \langle T^*\tilde{\omega}, Q - \widehat{Q} \rangle = 0\}$ separates the convex set $T^{-1}C$ and the convex level set of I which is "tangent" to $T^{-1}C$. Informally, the gradient of I at $\widehat{Q}$ is a normal vector of $T^{-1}C$ at $\widehat{Q}$: it must be a multiple of $T^*\tilde{\omega}$. Expressing this orthogonality relation by means of the convex conjugation (2.13) leads to the representation formula (2.15)(c). Clearly, Hahn-Banach theorem has something to do with the existence of $T^*\tilde{\omega}$. This picture is a guide for the proof in [19].

Following the terminology of Ney [22,23], as it shares the properties (2.15)(a,b) and (2.16), the minimizer $\hat{x}$ is called a *dominating point* of C with respect to Γ^* (see Def. 6.1 below).

2.9. Solving $(\overline{\mathrm{P}}_C)$

In this section, we study $(\overline{\mathrm{P}}_C)$ under the critical constraint hypothesis $(\mathrm{A}_\theta^\exists)$ which imposes that $T^*\mathcal{Y} \subset L_{\lambda_\diamond}$.

By Theorem 2.4, we have $L_\rho'' = [L_\rho \oplus L_{\rho^*}^s] \oplus L_\rho^{s\prime}$. For any $\zeta \in L_\rho'' = (L_{\rho^*} R \oplus L_\rho^s)'$, let us denote the restrictions $\zeta_1 = \zeta_{|L_{\rho^*} R}$ and $\zeta_2 = \zeta_{|L_\rho^s}$. Since, $(L_{\rho^*} R)' \simeq L_\rho \oplus L_{\rho^*}^s$, we see that any $\zeta \in L_\rho''$ is uniquely decomposed into

$$\zeta = \zeta_1^a + \zeta_1^s + \zeta_2 \tag{2.17}$$

with $\zeta_1 = \zeta_1^a + \zeta_1^s \in L_{\rho^*}'$, $\zeta_1^a \in L_\rho$, $\zeta_1^s \in L_{\rho^*}^s$ and $\zeta_2 \in L_\rho^{s\prime}$.

The extended dual problem $(\overline{D}_C)$. The extended dual problem associated with $(\overline{P}_C)$ is

$$maximize \quad \inf_{x \in C \cap \mathcal{X}} \langle \omega, x \rangle - I_\gamma\big([T^*\omega]_1^a\big), \quad \omega \in \overline{\mathcal{Y}} \qquad (\overline{D}_C)$$

where $\overline{\mathcal{Y}} \subset \mathcal{X}^*$ is some cone which contains $\mathcal{Y}_o$. Its exact definition is given at Appendix A.

Theorem 2.8. *Suppose that*

(1) the hypotheses (A) *are satisfied,*
(2) the convex set C is assumed to be such that

$$T^{-1}C \cap L'_{\lambda_\diamond} = \bigcap_{y \in Y} \big\{ \ell \in L'_{\lambda_\diamond}; \langle\langle y, \theta \rangle, \ell \rangle \geq a_y \big\} \qquad (2.18)$$

for some subset $Y \subset \mathcal{X}^*_o$ with $\langle y, \theta \rangle \in L_{\lambda_\diamond}$ for all $y \in Y$ and some function $y \in Y \mapsto a_y \in \mathbb{R}$. In other words, $T^{-1}C$ is a $\sigma(L'_{\lambda_\diamond}, L_{\lambda_\diamond})$-closed convex subset of $L'_{\lambda_\diamond}$.

Then:

(a) The dual equality for $(\overline{P}_C)$ is

$$\inf(\overline{P}_C) = \inf_{x \in C} \Gamma^*(x) = \sup(D_C) = \sup(\overline{D}_C) \in [0, \infty].$$

(b) If $C \cap \operatorname{dom}\Gamma^* \neq \emptyset$ or equivalently $C \cap T\operatorname{dom}\bar{I} \neq \emptyset$, then $(\overline{P}_C)$ admits solutions in $L'_{\lambda_\diamond}$, any minimizing sequence admits $\sigma(L'_{\lambda_\diamond}, L_{\lambda_\diamond})$-cluster points and every such point is a solution to $(\overline{P}_C)$.

    Suppose that in addition we have $C \cap \operatorname{icordom}\Gamma^* \neq \emptyset$ or equivalently $C \cap \operatorname{icor}(T\operatorname{dom}\bar{I}) \neq \emptyset$. Then:

(c) Let $\hat{\ell} \in L'_{\lambda_\diamond}$ be any solution to $(\overline{P}_C)$ and denote $\hat{x} := T\hat{\ell}$. There exists $\bar{\omega} \in \overline{\mathcal{Y}}$ such that

$$\begin{cases} (a) & \hat{x} \in C \cap \operatorname{dom}\Gamma^* \\ (b) & \langle \bar{\omega}, \hat{x} \rangle_{\mathcal{X}^*, \mathcal{X}} \leq \langle \bar{\omega}, x \rangle_{\mathcal{X}^*, \mathcal{X}}, \forall x \in C \cap \operatorname{dom}\Gamma^* \\ (c) & \hat{\ell} \in \gamma'_z([T^*\bar{\omega}]_1^a) R + D^\perp([T^*\bar{\omega}]_2) \end{cases} \qquad (2.19)$$

where we used notation (2.17) and $D^\perp([T^*\bar{\omega}]_2)$ is some cone in $L^s_{\lambda_\diamond}$ which is pointed at zero and whose direction depends on $[T^*\bar{\omega}]_2$ (see Appendix A for its precise definition).

    There exists some $\tilde{\omega} \in \mathcal{X}^*_o$ such that

$$[T^*\bar{\omega}]_1^a = \langle \tilde{\omega}, \theta(\cdot) \rangle_{\mathcal{X}^*_o, \mathcal{X}_o}$$

is a measurable function which can be approximated in some sense (see Appendix A) by sequences $(\langle y_n, \theta(\cdot) \rangle)_{n \geq 1}$ with $y_n \in \mathcal{Y}_o$ and $\int_{\mathcal{Z}} \lambda(\langle y_n, \theta \rangle) dR < \infty$ for all n.

    Furthermore, $\hat{\ell} \in L'_{\lambda_\diamond}$ and $\bar{\omega} \in \overline{\mathcal{Y}}$ satisfy (2.19) if and only if $\hat{\ell}$ solves $(\overline{P}_C)$ and $\bar{\omega}$ solves $(\overline{D}_C)$.

(d) Of course, (2.19)c implies $\hat{x} = \int_{\mathcal{Z}} \theta\gamma'(\langle \tilde{\omega}, \theta \rangle) dR + \langle \theta, \hat{\ell}^s \rangle$. Moreover,

    1. $\hat{x}$ minimizes Γ^* on C,
    2. $\bar{I}(\hat{\ell}) = \Gamma^*(\hat{x}) = \int_{\mathcal{Z}} \gamma^* \circ \gamma'(\langle \tilde{\omega}, \theta \rangle) dR + \sup\{\langle u, \hat{\ell}^s \rangle; u \in \operatorname{dom}I_\gamma\} < \infty$ and
    3. $\bar{I}(\hat{\ell}) + \int_{\mathcal{Z}} \gamma(\langle \tilde{\omega}, \theta \rangle) dR = \int_{\mathcal{Z}} \langle \tilde{\omega}, \theta \rangle d\hat{\ell}^a + \langle [T^*\bar{\omega}]_2, \hat{\ell}^s \rangle.$

Proof. This result is [19] (Thm. 4.2). $\qquad\qquad\square$

    The exact definitions of $\widetilde{\mathcal{Y}}$, $\overline{\mathcal{Y}}$ and $D^\perp$ as well as the precise statement of Theorem 2.8(c) are given at Appendix A. In particular, the complete statement of Theorem 2.8(c) is given at Theorem A.1.

    Figure 2 illustrates (2.19) which, with $\hat{x} = T\hat{\ell}$ can be rewritten: $\hat{\ell} \in T^{-1}C$ and $\langle T^*\bar{\omega}, \ell - \hat{\ell} \rangle \geq 0, \forall \ell \in T^{-1}C$. As in Figure 1, one sees that the hyperplane $\{\ell : \langle T^*\bar{\omega}, \ell - \hat{\ell} \rangle = 0\}$ separates the convex set $T^{-1}C$ and the convex level set of $\bar{I}$ which is "tangent" to $T^{-1}C$: the shaded area. The same kind of conclusions follow.
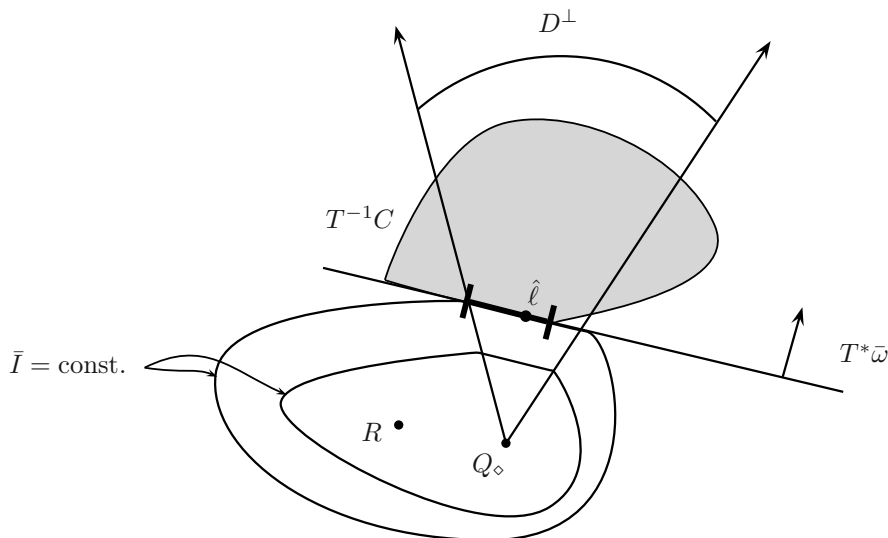
FIGURE 2.

The measure $Q_\diamond = \hat{\ell}^a = \gamma'_z([T^*\bar{\omega}]_1^a)\, R$ is the absolutely continuous part of $\hat{\ell}$; its expression is given at (4.1) below. The set of solutions $\hat{\ell}$ of $(\overline{\mathbf{P}}_C)$ is represented by the bold type segment, they all share the same absolutely continuous part $Q_\diamond$. The relation (2.19)(c) can be rewritten: $\hat{\ell} \in Q_\diamond + D^\perp$ which is the convex cone with vertex $Q_\diamond$ and direction $D^\perp$. The cone $D^\perp$ only contains singular directions. One sees with (2.9) and (2.10) that $D^\perp$ contributes to the positively homogeneous part I^s of $\bar{I}$. This is the reason why we decided to draw a flat part for the level lines of $\bar{I}$ when they cut $Q_\diamond + D^\perp$.

Figure 2 is only a guide for illustrating Theorem 2.8. It shouldn't be taken too seriously. For instance a better finite dimensional analogue would have been given by $\bar{I}(x, y, z) = x^2 + |y + z|$ which requires a 3D graphical representation.

3. SOME EXAMPLES

3.1. **Some examples of entropies**

Important examples of entropies occur in statistical physics, probability theory and mathematical statistics. Among them the relative entropy plays a prominent role.

Relative entropy

The reference measure R is assumed to be a probability measure. The relative entropy of $Q \in M_{\mathcal{Z}}$ with respect to $R \in P_{\mathcal{Z}}$ is

$$I(Q|R) = \begin{cases} \int_{\mathcal{Z}} \log\left(\frac{\mathrm{d}Q}{\mathrm{d}R}\right)\,\mathrm{d}Q & \text{if } Q \prec R \text{ and } Q \in P_{\mathcal{Z}} \\ +\infty & \text{otherwise,} \end{cases} \qquad Q \in M_{\mathcal{Z}}. \tag{3.1}$$

It corresponds to $\gamma_z^*(t) = \begin{cases} t\log t - t + 1 & \text{if } t > 0 \\ 1 & \text{if } t = 0 \\ +\infty & \text{if } t < 0, \end{cases} \quad m(z) = 1$ and

$$\lambda_z(s) = \mathrm{e}^s - s - 1, \quad s \in \mathbb{R}, z \in \mathcal{Z}. \tag{3.2}$$

A variant

Taking the same γ^* and removing the unit mass constraint gives

$$H(Q|R) = \begin{cases} \int_{\mathcal{Z}} \left[\frac{\mathrm{d}Q}{\mathrm{d}R} \log\left(\frac{\mathrm{d}Q}{\mathrm{d}R}\right) - \frac{\mathrm{d}Q}{\mathrm{d}R} + 1 \right] \mathrm{d}R & \text{if } 0 \leq Q \prec R \\ +\infty, & \text{otherwise,} \end{cases} \qquad Q \in M_{\mathcal{Z}}.$$

This entropy is the rate function of (1.4) when $(W_i)_{i \geq 1}$ is an iid sequence of Poisson (1) random weights. If R is σ-finite, it is the rate function of the LDP of normalized Poisson random measures, see [15].

Extended relative entropy

Since $\lambda(s) = \mathrm{e}^s - s - 1$ and $R \in P_{\mathcal{Z}}$ is a bounded measure, we have $\lambda_\diamond(s) = \tau(s) := \mathrm{e}^{|s|} - |s| - 1$ and the relevant Orlicz spaces are

$$L_{\tau^*} = \left\{ f : \mathcal{Z} \to \mathbb{R}; \int_{\mathcal{Z}} |f| \log|f| \, \mathrm{d}R < \infty \right\}$$

$$E_\tau = \left\{ u : \mathcal{Z} \to \mathbb{R}; \forall \alpha > 0, \int_{\mathcal{Z}} \mathrm{e}^{\alpha|u|} \, \mathrm{d}R < \infty \right\}$$

$$L_\tau = \left\{ u : \mathcal{Z} \to \mathbb{R}; \exists \alpha > 0, \int_{\mathcal{Z}} \mathrm{e}^{\alpha|u|} \, \mathrm{d}R < \infty \right\}$$

since $\tau^*(t) = (|t| + 1) \log(|t| + 1) - |t|$. The extended relative entropy is defined by

$$\bar{I}(\ell|R) = I(\ell^a|R) + \sup \left\{ \langle \ell^s, u \rangle; u, \int_{\mathcal{Z}} \mathrm{e}^u \, \mathrm{d}R < \infty \right\}, \quad \ell \in O_{\exp} \tag{3.3}$$

where $\ell = \ell^a + \ell^s$ is the decomposition into absolutely continuous and singular parts of ℓ in $L'_\tau = L_{\tau^*} \oplus L^s_\tau$, and

$$O_{\exp} = \{\ell \in L'_\tau; \ell \geq 0, \langle \ell, \mathbf{1} \rangle = 1\}. \tag{3.4}$$

Note that $O_{\exp}$ depends on R and that for all $\ell \in O_{\exp}$, $\ell^a \in P_{\mathcal{Z}} \cap L_{\tau^*} R$.

3.2. Some examples of constraints

Let us consider the two standard constraints which are the moment constraints and the marginal constraints.

Moment constraints

Let $\theta = (\theta_k)_{1 \leq k \leq K}$ be a measurable function from $\mathcal{Z}$ to $\mathcal{X}_o = \mathbb{R}^K$. The moment constraint is specified by the operator

$$\int_{\mathcal{Z}} \theta \, \mathrm{d}\ell = \left(\int_{\mathcal{Z}} \theta_k \, \mathrm{d}\ell \right)_{1 \leq k \leq K} \in \mathbb{R}^K,$$

which is defined for each $\ell \in M_{\mathcal{Z}}$ which integrates all the real valued measurable functions θ_k. The adjoint operator is

$$T^* y = \langle y, \theta \rangle = \sum_{1 \leq k \leq K} y_k \theta_k, \quad y = (y_1, \dots, y_K) \in \mathbb{R}^K.$$

Marginal constraints

Let $\mathcal{Z} = A \times B$ be a product space, M_{AB} be the space of all *bounded* signed measures on $A \times B$ and U_{AB} be the space of all measurable bounded functions u on $A \times B$. Denote $\ell_A = \ell(\cdot \times B)$ and $\ell_B = \ell(A \times \cdot)$ the marginal measures of $\ell \in M_{AB}$. The constraint of prescribed marginal measures is specified by

$$\int_{A \times B} \theta \, \mathrm{d}\ell = (\ell_A, \ell_B) \in M_A \times M_B, \quad \ell \in M_{AB}$$

where M_A and M_B are the spaces of all bounded signed measures on A and B. The function θ which gives the marginal constraint is

$$\theta(a, b) = (\delta_a, \delta_b), \ a \in A, b \in B$$

where δ_a is the Dirac measure at a. Indeed, $(\ell_A, \ell_B) = \int_{A \times B} (\delta_a, \delta_b) \, \ell(\mathrm{d}a \mathrm{d}b)$.

More precisely, let U_A, U_B be the spaces of measurable functions on A and B and take $\mathcal{Y}_o = U_A \times U_B$ and $\mathcal{X}_o = U_A^* \times U_B^*$. Then, θ is a measurable function from $\mathcal{Z} = A \times B$ to $\mathcal{X}_o = U_A^* \times U_B^*$ and the adjoint of the marginal operator

$$T\ell = \langle \theta, \ell \rangle = (\ell_A, \ell_B) \in U_A^* \times U_B^*, \quad \ell \in U_{AB}^*$$

where $\langle f, \ell_A \rangle := \langle f \otimes 1, \ell \rangle$ and $\langle g, \ell_B \rangle := \langle 1 \otimes g, \ell \rangle$ for all $f \in U_A$ and $g \in U_B$, is given by

$$T^*(f, g) = \langle (f, g), \theta \rangle = f \oplus g \in U_{AB}, \quad f \in U_A, g \in U_B \tag{3.5}$$

where $f \oplus g(a, b) := f(a) + g(b), \ a \in A, b \in B$.

4. Minimizing sequences

In this section, the minimization problem (P_C) is considered when the constraint function θ satisfies ($\mathrm{A}_\theta^\exists$) but not necessarily ($\mathrm{A}_\theta^\forall$). This means that the constraint is *critical*. Problem (P_C) may not be attained anymore. Nevertheless, minimizing sequences may admit a limit in some sense. As will be seen at Section 5, this phenomenon is tightly linked to the notion of *generalized entropic projection* introduced by Csiszár.

4.1. **Statements of the results**

We start this section stating its main results at Theorems 4.1 and 4.4.

Theorem 4.1 (attainment). *The hypotheses of Theorem 2.8 are assumed.*

(a) *Suppose that $C \cap \mathrm{dom}\,\Gamma^* \neq \emptyset$. Then, the minimization problem ($\overline{\mathrm{P}}_C$) is attained in $L'_{\lambda_\diamond}$ and all its solutions share the same absolutely continuous part $Q_\diamond \in L_{\lambda_\diamond^*} R$.*

(b) *Suppose that $C \cap \mathrm{icordom}\,\Gamma^* \neq \emptyset$. Then, ($\widetilde{\mathrm{D}}_C$) is attained in $\widetilde{\mathcal{Y}}$ and*

$$Q_\diamond(\mathrm{d}z) = \gamma_z'(\langle \omega_\diamond, \theta(z) \rangle) \, R(\mathrm{d}z) \tag{4.1}$$

where $\omega_\diamond \in \widetilde{\mathcal{Y}}$ is any solution to ($\widetilde{\mathrm{D}}_C$).

(c) *See also an additional statement at Proposition A.2.*

Proof.

• *Proof of (a).* The attainment statement is Theorem 2.8b. Let us show that as γ^* is strictly convex, if k_* and ℓ_* are two solutions of ($\overline{\mathrm{P}}_C$), their absolutely continuous parts match:

$$k_*^a = \ell_*^a. \tag{4.2}$$

k_*, ℓ_* are in the convex set $\{\ell \in L'_{\lambda_\diamond}; T\ell \in C\}$ and $\inf(\overline{P}_C) = \bar{I}(k_*) = \bar{I}(\ell_*)$. For all $0 \leq p, q \leq 1$ such that $p + q = 1$, as I and I^s are convex functions, we have

$$
\begin{aligned}
\inf(\overline{P}_C) &\leq \bar{I}(pk_* + q\ell_*) \\
&= I(pk_*^a + q\ell_*^a) + I^s(pk_*^s + q\ell_*^s) \\
&\leq pI(k_*^a) + qI(\ell_*^a) + pI^s(k_*^s) + qI^s(\ell_*^s) \\
&= p\bar{I}(k_*) + q\bar{I}(\ell_*) = \inf(\overline{P}_C).
\end{aligned}
$$

It follows that $I(pk_*^a + q\ell_*^a) + I^s(pk_*^s + q\ell_*^s) = pI(k_*^a) + qI(\ell_*^a) + pI^s(k_*^s) + qI^s(\ell_*^s)$. Suppose that $k_*^a \neq \ell_*^a$. As I is strictly convex, with $0 < p < 1$, we get: $I(pk_*^a + q\ell_*^a) < pI(k_*^a) + qI(\ell_*^a)$ and this implies that $I^s(pk_*^s + q\ell_*^s) > pI^s(k_*^s) + qI^s(\ell_*^s)$ which is impossible since I^s is convex. This proves (4.2).

• *Proof of (b).* Let $\bar{\ell}$ be any solution to $(\overline{P}_C)$. Denoting $\bar{x}^a = T\bar{\ell}^a$ and $\bar{x}^s = T\bar{\ell}^s$ we see with (2.19) that

$$
\begin{cases}
(a) & \bar{x}^a \in [C - \bar{x}^s] \cap \operatorname{dom}\Gamma^* \\
(b) & \langle \bar{\omega}, \bar{x}^a \rangle \leq \langle \bar{\omega}, x \rangle, \forall x \in [C - \bar{x}^s] \cap \operatorname{dom}\Gamma^* \\
(c) & \bar{\ell}^a = \gamma'_z(\langle \tilde{\omega}, \theta \rangle) R.
\end{cases}
$$

By Theorem 2.7c, this implies that $\tilde{\omega}$ solves $(\widetilde{D}_{C-\bar{x}^s})$.

It remains to show that $\tilde{\omega}$ also solves $(\widetilde{D}_C)$. Thanks to Theorem 2.8, we have: $\inf_{x \in C}\langle \bar{\omega}, x \rangle = \langle T^*\bar{\omega}, \bar{\ell} \rangle = \langle \langle \tilde{\omega}, \theta \rangle, \bar{\ell}^a \rangle + \bar{I}(\bar{\ell}^s)$ and $\inf_{x \in C}\langle \bar{\omega}, x \rangle - I_\gamma(\langle \tilde{\omega}, \theta \rangle) = \sup(\widetilde{D}_C) = \inf(\overline{P}_C) = I(\bar{\ell}^a) + \bar{I}(\bar{\ell}^s) = \inf(\overline{P}_{C-\bar{x}^s}) + \bar{I}(\bar{\ell}^s) = \sup(\widetilde{D}_{C-\bar{x}^s}) + \bar{I}(\bar{\ell}^s)$. Therefore $\inf_{x \in C}\langle \bar{\omega}, x \rangle = \inf_{x \in C-\bar{x}^s}\langle \tilde{\omega}, x \rangle + \bar{I}(\bar{\ell}^s)$ and subtracting $\bar{I}(\bar{\ell}^s)$ from $\sup(\widetilde{D}_C)$, we see that $\tilde{\omega}$ which solves $(\widetilde{D}_{C-\bar{x}^s})$ also solves $(\widetilde{D}_C)$. We complete the proof of the theorem, taking $\omega_\diamond = \tilde{\omega}$.

□

Remark 4.2. Replacing $\bar{\ell}^s$ with $t\bar{\ell}^s$, the same proof shows that $\tilde{\omega}$ solves $(\widetilde{D}_{C+(t-1)\bar{x}^s})$ for any $t \geq 0$.

From now on, we denote $Q_\diamond \in L_{\lambda_\diamond^*}R$ the absolutely continuous part shared by all the solutions of $(\overline{P}_C)$. Let us introduce

$$
\begin{aligned}
\mathcal{C} &= \left\{ Q \in L_{\lambda_\diamond^*}R; \ TQ := \int_{\mathcal{Z}} \theta \, dQ \in C \right\} \\
\overline{\mathcal{C}} &= \left\{ \ell \in L'_{\lambda_\diamond}; \ T\ell := \langle \theta, \ell \rangle \in C \right\}
\end{aligned}
\tag{4.3}
$$

the constraint sets $T^{-1}C \cap L_{\lambda_\diamond^*}R$ and $T^{-1}C$ on which I and $\bar{I}$ are minimized. We have: $\mathcal{C} = \overline{\mathcal{C}} \cap L_{\lambda_\diamond^*}R$ and $I = \bar{I} + \iota_{L_{\lambda_\diamond^*}R}$. Hence, $\inf(\overline{P}_C) \leq \inf(P_C)$. Clearly, $\mathcal{C} \cap \operatorname{dom} I \neq \emptyset \Leftrightarrow \inf(P_C) < \infty$ implies $\overline{\mathcal{C}} \cap \operatorname{dom} \bar{I} \neq \emptyset \Leftrightarrow \inf(\overline{P}_C) < \infty \Leftrightarrow C \cap \operatorname{dom}\Gamma^* \neq \emptyset$.

Of course, if $\mathcal{C} \cap \operatorname{dom} I \neq \emptyset$, (P_C) admits nontrivial minimizing sequences. Theorem 4.4 below gives some details about them.

The present paper is concerned with

$$
minimize \ I(Q) \ subject \ to \ Q \in \mathcal{C}. \tag{$\mathcal{P}_\mathcal{C}$}
$$

where $\mathcal{C}$ has the form (4.3). But this is not a restriction as explained in the following remarks.

Remark 4.3.

(1) Taking T^* to be the identity on $\mathcal{Y}_o = E_{\lambda_\diamond}$ or $L_{\lambda_\diamond}$ (being careless with a.e. equality, this corresponds to $\theta(z)$ to be the Dirac measure δ_z), we see that $(A_\theta^\vee)$ or $(A_\theta^\exists)$ is satisfied respectively. Hence, with $C = \mathcal{C}$, (P_C) is $(\mathcal{P}_\mathcal{C})$. Consequently, the specific form with θ and C adds details to the description of $\mathcal{C}$ without loss of generality.

(2) With θ, $\mathcal{Y}_o$ and C as in (1), the assumptions on $C = \mathcal{C}$ are:
    (a) Under $(A_\theta^\vee)$, (2.14) is equivalent to $\mathcal{C}$ is $\sigma(L_{\lambda_\diamond^*}R, E_{\lambda_\diamond})$-closed.
        Note that if $\lambda_\diamond$ and $\lambda_\diamond^*$ both satisfy the Δ_2-condition, as $\mathcal{C}$ is convex, this is equivalent to $\mathcal{C}$ being $\| \cdot \|_{\lambda_\diamond^*}$-closed.
    (b) Under $(A_\theta^\exists)$, (2.18) is equivalent to $\mathcal{C}$ is $\sigma(L_{\lambda_\diamond^*}R, L_{\lambda_\diamond})$-closed.
        Note that if $\lambda_\diamond^*$ satisfies the Δ_2-condition, as $\mathcal{C}$ is convex, this is equivalent to $\mathcal{C}$ being $\| \cdot \|_{\lambda_\diamond^*}$-closed.

We denote $\| \cdot \|_{\lambda_\diamond^*}$-int $(\mathcal{C})$ the interior of $\mathcal{C}$ in $L_{\lambda_\diamond^*}R$ with respect to the strong topology of $L_{\lambda_\diamond^*}$.

Theorem 4.4 (minimizing sequences of (P_C)). *Assume that the hypotheses* (A) *hold,*

$$\lim_{t \to \pm\infty} \gamma_z^*(t)/t = +\infty, \quad \text{for } R\text{-almost every } z \in \mathcal{Z} \tag{4.4}$$

and $\mathcal{C}$ which is defined at (4.3) *is $\sigma(L_{\lambda_\diamond^*}R, L_{\lambda_\diamond})$-closed and satisfies $\mathcal{C} \cap \operatorname{dom} I \neq \emptyset$.*
    Let us consider the following additional conditions.
    (1) (a) There are finitely many moment constraints, i.e. $\mathcal{X}_o = \mathbb{R}^K$ (see Sect. 3.2);
        (b) $\mathcal{C} \cap \operatorname{icordom} I \neq \emptyset$.
    (2) $\mathcal{C}$ has a nonempty $\| \cdot \|_{\lambda_\diamond^*}$-interior.
Under one of these additional conditions (1) *or* (2), *we have*

$$I(Q_\diamond) \leq \inf(P_C) = \inf(\overline{P}_C)$$

and any minimizing sequence of (P_C) converges to $Q_\diamond$ with respect to $\sigma(L_{\lambda_\diamond^*}R, E_{\lambda_\diamond})$ and in variation norm (i.e. strongly in L_1R).

Proof. This proof relies on results which are stated and proved in the remainder of the present section. It is shown at Lemma 4.6 that any minimizing sequence of (P_C) converges in the sense of the $\sigma(L_{\lambda_\diamond^*}R, E_{\lambda_\diamond})$-topology to $Q_\diamond$, whenever $\inf(P_C) = \inf(\overline{P}_C)$. But this equality holds thanks to Lemma 4.11 and
- under condition (1): Lemma 4.12a;
- under condition (2): Corollary 4.14b.

Let us have a look at the last inequality. For any $\bar{\ell} = \bar{\ell}^a + \bar{\ell}^s = Q_\diamond + \bar{\ell}^s$ minimizer of $(\overline{P}_C)$ and any $(Q_n)_{n \geq 1}$ minimizing sequence of (P_C), we obtain

$$\begin{aligned}
\inf_n I(Q_n) &= \inf(P_C) = \inf(\overline{P}_C) = \bar{I}(\bar{\ell}) \\
&= I(Q_\diamond) + I^s(\bar{\ell}^s) \\
&\geq I(Q_\diamond)
\end{aligned}$$

with a strict inequality if $I^s(\bar{\ell}^s) > 0$.
    Finally, (4.4) implies that $\lambda_\diamond$ is finite. As R is assumed to be bounded, we have $L_\infty \subset E_{\lambda_\diamond}$. But, we also have $L_{\lambda_\diamond^*}R \subset L_1R$ so that the $\sigma(L_{\lambda_\diamond^*}R, L_{\lambda_\diamond})$-convergence implies the $\sigma(L_1R, L_\infty)$-convergence of any minimizing sequence. The strong convergence in L_1R now follows from [2] (Thm. 3.7). $\qquad\square$

Remark 4.5.
    (1) As regards condition (1), it is not assumed that $\mathcal{C}$ has a nonempty interior.
    (2) As regards condition (2):
        (i) Any $\sigma(L_{\lambda_\diamond^*}R, L_{\lambda_\diamond})$-closed convex set has the form

$$\mathcal{C} = \bigcap_{u \in U} \left\{ \ell \in L_{\lambda_\diamond^*}R; \langle u, \ell \rangle \geq a_u \right\}$$

        for some $U \subset L_{\lambda_\diamond}$.

(ii) For $\mathcal{C}$ to have a nonempty $\|\cdot\|_{\lambda_\diamond^*}$-interior, it is enough that $C \cap \mathcal{X}$ has a nonempty interior in $\mathcal{X}$ endowed with the dual norm $|\cdot|_{\lambda_\diamond}^*$ defined at (4.8). This is a consequence of Lemma 4.9(a) below.

(3) The last quantity $I^s(\bar{\ell}^s) = \inf(\mathrm{P}_C) - I(Q_\diamond)$ is precisely the *gap* of lower $\sigma(L_{\lambda_\diamond^*}R, E_{\lambda_\diamond})$-semicontinuity of I : $\lim_n Q_n = Q_\diamond$ and $\inf(\mathrm{P}_C) = \liminf_n I(Q_n) \geq I(\lim_n Q_n) = I(Q_\diamond)$.

4.2. A preliminary lemma

Preliminary results for the proof of Theorem 4.4 are stated below at Lemma 4.6. The assumption $(\mathrm{A}_\theta^\exists)$ about the critical constraint is $T^* \mathcal{Y}_o \subset L_{\lambda_\diamond}$.

Lemma 4.6. *Assume that the hypotheses* (A) *and* (4.4) *hold,* $\mathcal{C}$ *is* $\sigma(L_{\lambda_\diamond^*}R, L_{\lambda_\diamond})$-*closed,* $\inf(\mathrm{P}_C) < \infty$ *and*

$$\inf(\mathrm{P}_C) = \inf(\overline{\mathrm{P}}_C). \tag{4.5}$$

Then, any minimizing sequence of (P_C) *converges to* $Q_\diamond$ *with respect to* $\sigma(L_{\lambda_\diamond^*}R, E_{\lambda_\diamond})$.

Proof. Let $(Q_n)_{n \geq 1}$ be a minimizing sequence of (P_C). Since it is assumed that $\inf(\mathrm{P}_C) = \inf(\overline{\mathrm{P}}_C)$, $(Q_n)_{n \geq 1}$ is also a minimizing sequence of $(\overline{\mathrm{P}}_C)$. By [19] (Lem. 6.2), $\bar{I}$ is $\sigma(L'_{\lambda_\diamond}, L_{\lambda_\diamond})$-inf-compact. Hence, we can extract a $\sigma(L'_{\lambda_\diamond}, L_{\lambda_\diamond})$-convergent subnet $(Q_\alpha)_\alpha$ from $(Q_n)_{n \geq 1}$. Let $\ell_* \in \overline{\mathcal{C}}$ denote its limit: we have $\lim_\alpha \int_{\mathcal{Z}} u \, dQ_\alpha = \langle \ell_*, u \rangle$ for all $u \in L_{\lambda_\diamond}$. As $\langle \ell_*^s, u \rangle = 0$, for all $u \in E_{\lambda_\diamond}$ (see Prop. 2.3), we obtain: $\lim_\alpha \int_{\mathcal{Z}} u \, dQ_\alpha = \int_{\mathcal{Z}} u \, d\ell_*^a$ for all $u \in E_{\lambda_\diamond}$. This proves that $(Q_\alpha)_\alpha$ $\sigma(E'_{\lambda_\diamond}, E_{\lambda_\diamond})$-converges to ℓ_*^a. As $E_{\lambda_\diamond}$ is a separable Banach space ($L_{\lambda_\diamond}$ is not separable in general), the topology $\sigma(E'_{\lambda_\diamond}, E_{\lambda_\diamond}) = \sigma(L_{\lambda_\diamond^*}R, E_{\lambda_\diamond})$ is metrizable and we can extract a convergent sub*sequence* $(\tilde{Q}_k)_{k \geq 1}$ from the convergent net $(Q_\alpha)_\alpha$. Hence, $(\tilde{Q}_k)_{k \geq 1}$ is a subsequence of $(Q_n)_{n \geq 1}$ which $\sigma(L_{\lambda_\diamond^*}R, E_{\lambda_\diamond})$-converges to ℓ_*^a.

Since $\bar{I}$ is inf-compact, ℓ_* is a minimizer of $(\overline{\mathrm{P}}_C)$ and by Theorem 4.1a, there is a unique $Q_\diamond$ such for any minimizing sequence $(Q_n)_{n \geq 1}$, $\ell_*^a = Q_\diamond$. Therefore, any convergent subsequence of $(Q_n)_{n \geq 1}$ converges to $Q_\diamond$ for $\sigma(L_{\lambda_\diamond^*}R, E_{\lambda_\diamond})$. As any subsequence of a minimizing sequence is still a minimizing sequence, we have proved that from any subsequence of $(Q_n)_{n \geq 1}$, we can extract a sub-subsequence which converges to $Q_\diamond$. This proves that $(Q_n)_{n \geq 1}$ converges to $Q_\diamond$ with respect to $\sigma(L_{\lambda_\diamond^*}R, E_{\lambda_\diamond})$. $\qquad\square$

4.3. Sufficient conditions for $\inf(\mathrm{P}_C) = \inf(\overline{\mathrm{P}}_C)$

Our aim now is to obtain sufficient conditions for the identity $\inf(\mathrm{P}_C) = \inf(\overline{\mathrm{P}}_C)$ to hold. Let us rewrite the problems (P_C) and $(\overline{\mathrm{P}}_C)$ in order to emphasize their differences and analogies. Denote

$$\begin{aligned}
\Phi_L(u) &= I_\lambda(u) = \int_{\mathcal{Z}} \lambda(u) \, \mathrm{d}R, \quad u \in L_{\lambda_\diamond} \\
\Phi_E(u) &= \Phi_L(u) + \iota_{E_{\lambda_\diamond}}(u), \quad u \in L_{\lambda_\diamond}
\end{aligned}$$

where E and L refer to $E_{\lambda_\diamond}$ and $L_{\lambda_\diamond}$. Their convex conjugates are

$$\begin{aligned}
\Phi_E^*(\ell) &= \sup_{u \in E_{\lambda_\diamond}} \{\langle \ell, u \rangle - I_\lambda(u)\}, \quad \ell \in L_{\lambda_\diamond^*}R \\
\Phi_L^*(\ell) &= \sup_{u \in L_{\lambda_\diamond}} \{\langle \ell, u \rangle - I_\lambda(u)\}, \quad \ell \in L'_{\lambda_\diamond}.
\end{aligned}$$

It is shown at [19] (Lem. 6.2) that under the assumption (4.4)

$$\begin{cases}
I(\ell) &= \Phi_E^*(\ell - mR), \quad \ell \in E'_{\lambda_\diamond} = L_{\lambda_\diamond^*}R \\
\bar{I}(\ell) &= \Phi_L^*(\ell - mR), \quad \ell \in L'_{\lambda_\diamond} = L_{\lambda_\diamond^*}R \oplus L^s_{\lambda_\diamond}.
\end{cases} \tag{4.6}$$

Hence, considering the minimization problems

$$minimize \ \Phi_E^*(\ell) \ subject \ to \ \langle\theta,\ell\rangle \in C_o, \quad \ell \in L_{\lambda_\diamond^*}R \tag{P_E}$$

and

$$minimize \ \Phi_L^*(\ell) \ subject \ to \ \langle\theta,\ell\rangle \in C_o, \quad \ell \in L'_{\lambda_\diamond} \tag{P_L}$$

with $C_o = C - \langle\theta, mR\rangle$, we see that ℓ_* is a solution of (P_C) [resp. $(\overline{P}_C)$] if and only if $\ell_* - mR$ is a solution of (P_E) [resp. (P_L)]. It is enough to prove

$$\inf(P_E) = \inf(P_L) \tag{4.7}$$

to get $\inf(P_C) = \inf(\overline{P}_C)$.

Basic facts about convex duality. The proof of (4.7) will rely on standard convex duality considerations. Let us recall some related facts, as developed in [26].

Let A and X be two vector spaces, $h : A \to [-\infty, +\infty]$ a convex function, $T : A \to X$ a linear operator and C a convex subset of X. The primal problem to be considered is the following convex minimization problem

$$minimize \ h(a) \ subject \ to \ Ta \in C, \quad a \in A. \tag{$\mathcal{P}$}$$

The primal value-function corresponding to the Fenchel perturbation $F(a, x) = h(a) + \iota_C(Ta + x)$, $a \in A, x \in X$ is $\varphi(x) = \inf_{a\in A} F(a, x)$, *i.e.*

$$\varphi(x) = \inf\{h(a); a \in A, Ta \in C - x\} \quad x \in X.$$

Denote A^* the algebraic dual space of A. The convex conjugate of h with respect to the dual pairing $\langle A, A^*\rangle$ is

$$h^*(\nu) = \sup_{a\in A}\{\langle\nu, a\rangle - h(a)\}, \quad \nu \in A^*.$$

Let Y be a vector space in dual pairing with X. The adjoint operator $T^* : Y \to A^*$ of T is defined for all $y \in Y$ by

$$\langle T^*y, a\rangle_{A^*,A} = \langle y, Ta\rangle_{Y,X}, \quad \forall a \in A.$$

The dual problem associated with $(\mathcal{P})$ is

$$maximize \ \inf_{x\in C}\langle y, x\rangle - h^*(T^*y), \quad y \in Y. \tag{$\mathcal{D}$}$$

Let U be some subspace of A^*. The dual value-function is

$$\psi(u) = \sup_{y\in Y}\left\{\inf_{x\in C}\langle y, x\rangle - h^*(T^*y + u)\right\}, \quad u \in U.$$

We say that $\langle X, Y\rangle$ is a *topological dual pairing* if X and Y are topological vector spaces and their topological dual spaces X' and Y' satisfy $X' = Y$ and $Y' = X$ up to some isomorphisms.

Theorem 4.7 (criteria for the dual equality). *We assume that A, U, X and Y are locally convex Hausdorff topological vector spaces such that $\langle A, U\rangle$ and $\langle X, Y\rangle$ are topological dual pairings. For the dual equality*

$$\inf(\mathcal{P}) = \sup(\mathcal{D})$$

to hold, it is enough that

(1) *(a) h is a convex function and C is a convex subset of X,*
     (b) φ is lower semicontinuous at $0 \in X$ and
     (c) $\sup(\mathcal{D}) > -\infty$

or

(2) *(a) h is a convex function and C is a closed convex subset of X;*
     (b) ψ is upper semicontinuous at $0 \in U$; and
     (c) $\inf(\mathcal{P}) < +\infty$.

Remark 4.8. About the space U.

(a) As regards criterion (1), the space U is unnecessary.

(b) As regards criterion (2), it is not assumed that $T^*Y \subset U$.

The continuity of T and T^*. The following Lemma 4.9 will be useful for the proof of (4.7). We go back to our usual notation. The Luxemburg norm on $L_{\lambda_\diamond}$ is $\| \cdot \|_{\lambda_\diamond}$ and its dual norm is

$$\|\ell\|_{\lambda_\diamond}^* := \sup_{u, \|u\|_{\lambda_\diamond} \leq 1} |\langle u, \ell \rangle|, \quad \ell \in L'_{\lambda_\diamond}.$$

Let us define

$$|y|_{\lambda_\diamond} = \|\langle y, \theta \rangle\|_{\lambda_\diamond}, \quad y \in \mathcal{Y}_o.$$

Under the assumption (A_θ), $|\cdot|_{\lambda_\diamond}$ is a norm on $\mathcal{Y}_o$. The dual space of $(\mathcal{Y}_o, |\cdot|_{\lambda_\diamond})$ is $\mathcal{X}$ and the corresponding dual norm is

$$|x|_{\lambda_\diamond}^* := \sup_{y, |y|_{\lambda_\diamond} \leq 1} |\langle y, x \rangle|, \quad x \in \mathcal{X}. \tag{4.8}$$

Lemma 4.9. *Let us assume (A_θ).*

(a) $T : L'_{\lambda_\diamond} \to \mathcal{X}$ is $\| \cdot \|_{\lambda_\diamond}^* - | \cdot |_{\lambda_\diamond}^*$-continuous,
(b) $T^*\mathcal{Y} \subset L_{\lambda_\diamond}$ and $T^* : \mathcal{Y} \to L_{\lambda_\diamond}$ is $\sigma(\mathcal{Y}, \mathcal{X})$-$\sigma(L_{\lambda_\diamond}, L'_{\lambda_\diamond})$-continuous,
(c) $T : L'_{\lambda_\diamond} \to \mathcal{X}$ is $\sigma(L'_{\lambda_\diamond}, L_{\lambda_\diamond})$-$\sigma(\mathcal{X}, \mathcal{Y})$-continuous.

Proof. See [16] (Sect. 4). $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Back to our problem. Let us particularize this framework for the problems (P_E) and (P_L). Assuming that $m \equiv 0$, we see that $(P_C)=(P_E)$, $(\overline{P}_C)=(P_L)$, $I = I_{\lambda^*} = \Phi_E^*$, $\bar{I} = \bar{I}_{\lambda^*} = \Phi_L^*$, $\gamma = \lambda$, $C = C_o$ and so on. *This simplifying requirement will be assumed during the proof without loss of generality, see the proof of [19] (Thm. 4.2.)*

Let us first apply the criterion (1) of Theorem 4.7.

Problem (P_E) is obtained with $A = L_{\lambda_\diamond^*}R$, $X = \mathcal{X}$, $Y = \mathcal{Y}$ equipped with the weak topologies $\sigma(\mathcal{X}, \mathcal{Y})$ and $\sigma(\mathcal{Y}, \mathcal{X})$, and $h = \Phi_E^* = I$. The corresponding primal value-function is

$$\varphi_E(x) = \inf \left\{ I(Q); \int_{\mathcal{Z}} \theta \, \mathrm{d}Q \in C - x, Q \in L_{\lambda_\diamond^*}R \right\}, \quad x \in \mathcal{X}.$$

Under the underlying assumption $(A_\theta^\exists)$, with Lemma 4.9(b) we have:

$$T^*\mathcal{Y} \subset L_{\lambda_\diamond}. \tag{4.9}$$

Hence, we only need to compute h^* on $L_{\lambda_\diamond} \subset [L_{\lambda_\diamond^*}R]^* = A^*$. For each $u \in L_{\lambda_\diamond}$, $h^*(u) = \sup_{f \in L_{\lambda_\diamond^*}} \{\int_{\mathcal{Z}} uf \, \mathrm{d}R - \int_{\mathcal{Z}} \lambda^*(f) \, \mathrm{d}R\}$ and it is proved in [25] that

$$h^*(u) = \int_{\mathcal{Z}} \lambda(u) \, \mathrm{d}R, \quad u \in L_{\lambda_\diamond}. \tag{4.10}$$

Therefore, the dual problem associated to (P_E) is

$$maximize \ \inf_{x \in C} \langle y, x \rangle - \int_{\mathcal{Z}} \lambda(T^*y) \, dR, \quad y \in \mathcal{Y}. \tag{D_E}$$

Let us go on with (P_L). Take $A = L'_{\lambda_\diamond}$, $X = \mathcal{X}$, $Y = \mathcal{Y}$ equipped with the weak topologies $\sigma(\mathcal{X}, \mathcal{Y})$ and $\sigma(\mathcal{Y}, \mathcal{X})$, and $h = \Phi_L^* = \bar{I}$. The function h^* in restriction to $L_{\lambda_\diamond}$ is still given by (4.10) since $u \in L_{\lambda_\diamond} \mapsto \int_{\mathcal{Z}} \lambda(u) \, dR$ is closed convex (Fatou's lemma). The primal value-function is

$$\varphi_L(x) = \inf\{\bar{I}(\ell); \langle \theta, \ell \rangle \in C - x, \ell \in L'_{\lambda_\diamond}\}, \quad x \in \mathcal{X}$$

and the dual problem associated to (P_L) is

$$(D_L) = (D_E).$$

Lemma 4.10. *Suppose that $T^*\mathcal{Y} \subset L_{\lambda_\diamond}$ and $C \cap \mathcal{X}$ is $\sigma(\mathcal{X}, \mathcal{Y})$-closed, then φ_L is $\sigma(\mathcal{X}, \mathcal{Y})$-lower semicontinuous.*

Proof. Defining $\tilde{\varphi}(x) := \varphi_L(-x)$ and $\bar{J}(x) := \inf\{\bar{I}(\ell); \ell \in L'_{\lambda_\diamond} : \langle \theta, \ell \rangle = x\}$, $x \in \mathcal{X}$, we obtain that $\tilde{\varphi}$ is the inf-convolution of $\bar{J}$ and the convex indicator of $-C : \iota_{-C}$. That is $\tilde{\varphi}(x) = (\bar{J} \square \iota_{-C})(x) = \inf\{\bar{J}(y) + \iota_{-C}(z); y, z, y + z = x\}$.

As already seen, $\bar{I}$ is $\sigma(L'_{\lambda_\diamond}, L_{\lambda_\diamond})$-inf-compact and T is $\sigma(L'_{\lambda_\diamond}, L_{\lambda_\diamond})$-$\sigma(\mathcal{X}, \mathcal{Y})$-continuous, see Lemma 4.9(c). It follows that $\bar{J}$ is $\sigma(\mathcal{X}, \mathcal{Y})$-inf-compact. As $C \cap \mathcal{X}$ is assumed to be $\sigma(\mathcal{X}, \mathcal{Y})$-closed, ι_{-C} is lower semicontinuous. Finally, being the inf-convolution of an inf-compact function and a lower semicontinuous function, $\tilde{\varphi}$ is lower semicontinuous, and so is φ_L. $\square$

As $\bar{I}$ and C are assumed to be convex and $\sup(D_L) \geq \inf_{x \in C} \langle 0, x \rangle - \int_{\mathcal{Z}} \lambda(T^*0) \, dR = 0 > -\infty$, this lemma allows us to apply criterion (1) of Theorem 4.7 to obtain

$$\inf(P_L) = \sup(D_L). \tag{4.11}$$

Since $\bar{I}$ and I match on $L_{\lambda_\diamond^*}R$, we have $\inf(P_L) \leq \inf(P_E)$. Putting together these considerations gives us

$$\sup(D_E) = \sup(D_L) = \inf(P_L) \leq \inf(P_E).$$

Since the desired equality (4.5) is equivalent to $\inf(P_L) = \inf(P_E)$, we have proved

Lemma 4.11. *The equality (4.5) holds if and only if we have the dual equality*

$$\inf(P_E) = \sup(D_E). \tag{4.12}$$

This happens if and only if φ_E is $\sigma(\mathcal{X}, \mathcal{Y})$-lower semicontinuous at $x = 0$.

Let us now give a couple of simple criteria for this property to be realized.

Lemma 4.12.

(a) Suppose that there are finitely many constraints (i.e. $\mathcal{X}_o$ is finite dimensional) and $\mathcal{C} \cap \mathrm{icordom}\, I \neq \emptyset$, then φ_E is continuous at 0.

(b) Suppose that $(\mathrm{A}_\theta^\vee)$ is satisfied and C is $\sigma(\mathcal{X}, \mathcal{Y})$-closed, then φ_E is $\sigma(\mathcal{X}, \mathcal{Y})$-lower semicontinuous.

Proof.

- *Proof of (a).* To get (a), remark that a convex function on a finite dimensional space is lower semicontinuous on the intrinsic core of its effective domain. By Lemma 4.9(c), T is $\sigma(L_{\lambda_\diamond^*}R, L_{\lambda_\diamond})$-continuous and the assumption $\mathcal{C} \cap \mathrm{icordom}\, I \neq \emptyset$ implies that 0 belongs to $\mathrm{icordom}\, \varphi_E$.

- *Proof of (b).* It is similar to the proof of Lemma 4.10. The assumption $(\mathrm{A}_\theta^\vee)$ insures that T is $\sigma(L_{\lambda_\diamond^*}R, E_{\lambda_\diamond})$-$\sigma(\mathcal{X}, \mathcal{Y})$-continuous.

$\square$

It follows from Lemmas 4.12b, 4.11, the remark at (4.7) and Lemma 4.6 that under the good constraint assumption $(\mathrm{A}_\theta^\vee)$, if $C \cap \mathrm{dom}\, \Gamma^* \neq \emptyset$, then any minimizing sequence of (P_C) converges with respect to the topology $\sigma(L_{\lambda_\diamond^*}R, E_{\lambda_\diamond})$ to the unique solution $\widehat{Q}$ of (P_C). This is Theorem 2.7b.

Using criterion (2). Up to now, we only used criterion (1) of Theorem 4.7. In the following lines, we are going to use criterion (2) to prove (4.12) under additional assumptions.

Let us go back to problem (P_E). It is still assumed without loss of generality that $m = 0$ and $\gamma = \lambda$. We introduce a space U and a dual value-function ψ on U. The framework of Theorem 4.7 is preserved when taking $X = \mathcal{X}$, $Y = \mathcal{Y}$ with the weak topologies $\sigma(\mathcal{X}, \mathcal{Y})$ and $\sigma(\mathcal{Y}, \mathcal{X})$, $h = I$ on $A = L_{\lambda_\diamond^*}R$ as before and adding the following topological pairing $\langle A, U \rangle$. We endow $A = L_{\lambda_\diamond^*}R$ with the topology $\sigma(L_{\lambda_\diamond^*}R, L_{\lambda_\diamond})$ and take

$$U = A' = (L_{\lambda_\diamond^*}R)' \simeq L_{\lambda_\diamond}$$

with the topology $\sigma(L_{\lambda_\diamond}, L_{\lambda_\diamond^*})$. By (4.10): $h^* = I_\lambda$, this leads to the dual value-function

$$\psi(u) = \sup_{y \in \mathcal{Y}} \left\{ \inf_{x \in C} \langle y, x \rangle - I_\lambda(T^*y + u) \right\}, \quad u \in L_{\lambda_\diamond}.$$

To apply criterion (2), let us establish the following

Lemma 4.13. *For ψ to be $\sigma(L_{\lambda_\diamond}, L_{\lambda_\diamond^*})$-upper semicontinuous, it is enough that*

(a) $T^*\mathcal{Y}$ is a $\sigma(L_{\lambda_\diamond}, L_{\lambda_\diamond^*})$-closed subspace of $L_{\lambda_\diamond}$ and

(b) the interior of $\mathcal{C}$ in $L_{\lambda_\diamond^*}R$ with respect to $\|\cdot\|_{\lambda_\diamond^*}$ is nonempty.

Proof. During this proof, unless specified the topology on $L_{\lambda_\diamond}$ is $\sigma(L_{\lambda_\diamond}, L_{\lambda_\diamond^*})$. For all $u \in L_{\lambda_\diamond}$,

$$
\begin{aligned}
-\psi(u) &= \inf_{y \in \mathcal{Y}} \left\{ \sup\{\langle -T^*y, \ell \rangle; \ell \in L_{\lambda_\diamond^*}R : T\ell \in C\} + I_\lambda(T^*y + u) \right\} \\
&= \inf_{v \in V} \left\{ \sup\{\langle -v, \ell \rangle; \ell \in \mathcal{C}\} + I_\lambda(v + u) \right\} \\
&= I_\lambda \square G(u)
\end{aligned}
$$

where $V = T^*\mathcal{Y}$ and $I_\lambda \square G(u) = \inf_{v \in L_{\lambda_\diamond}}\{G(v) + I_\lambda(u - v)\}$ is the inf-convolution of I_λ and $G(u) = \iota_{\mathcal{C}}^*(u) + \iota_V(u), u \in L_{\lambda_\diamond}$. Let us show that under the assumption (a),

$$G = \iota_{\mathcal{C}}^*.$$

As V is assumed to be closed, we have $\iota_V = \iota_{V^\perp}^*$ with $V^\perp = \{k \in L_{\lambda_\diamond^*}R; \langle v, k \rangle = 0, \forall v \in V\}$. This gives for each $u \in L'_{\lambda_\diamond^*}$, $G(u) = \iota_{\mathcal{C}}^*(u) + \iota_{V^\perp}^*(u) = \sup_{\ell \in \mathcal{C}}\langle u, \ell \rangle + \sup_{k \in V^\perp}\langle u, k \rangle = \sup\{\langle u, \ell + k \rangle; \ell \in \mathcal{C}, k \in V^\perp\} = \iota_{\mathcal{C}+V^\perp}^*(u) = \iota_{\mathcal{C}}^*(u)$, where the last equality holds since $\mathcal{C} + V^\perp = \mathcal{C}$, note that $\ker T = V^\perp$.

As convex conjugates, $\iota_{\mathcal{C}}^*$ and $I_\lambda = h^*$ are closed convex functions.

Since for all $u, v \in L_{\lambda_\diamond}$,

$$
\begin{aligned}
\iota_{\mathcal{C}}^*(v) + I_\lambda(u - v) &= \iota_{\mathcal{C}-k}^*(v) - \langle v, k \rangle + I_\lambda(u - v) \\
&= \iota_{\mathcal{C}-k}^*(v) + \langle u - v, k \rangle + I_\lambda(u - v) - \langle u, k \rangle,
\end{aligned}
$$

we have

$$
-\psi + \langle \cdot, k \rangle = \iota_{\mathcal{C}-k}^* \square (I_\lambda - \langle \cdot, k \rangle).
$$

But, by the assumption (b) there exists some k in $L_{\lambda_\diamond^*} R$ such that $0 \in \text{int}\,(\mathcal{C} - k)$. It follows that $\iota_{\mathcal{C}-k}^*$ is inf-compact. Finally, $-\psi + \langle \cdot, k \rangle$ is lower semicontinuous, being the inf-convolution of a lower semicontinuous and an inf-compact functions. $\qquad \square$

Corollary 4.14. *Assume that the hypotheses* (A) *and* (4.4) *hold.*

(a) *Assume in addition that C is $\sigma(\mathcal{X}, \mathcal{Y})$-closed convex, $T^*\mathcal{Y}$ is a $\sigma(L_{\lambda_\diamond}, L_{\lambda_\diamond^*})$-closed subspace of $L_{\lambda_\diamond}$, $\mathcal{C}$ has a nonempty $\|\cdot\|_{\lambda_\diamond^*}$-interior and $\inf(\mathrm{P}_E) < \infty$. Then, (4.12) is satisfied.*

(b) *In particular, if $\mathcal{C}$ is $\sigma(L_{\lambda_\diamond}R, L_{\lambda_\diamond})$-closed convex set with a nonempty $\|\cdot\|_{\lambda_\diamond^*}$-interior and $\inf(\mathrm{P}_E) < \infty$, then (4.12) is satisfied.*

Proof.

- *Proof of (a).* Apply the criterion (2) of Theorem 4.7 with Lemma 4.13.

- *Proof of (b).* This is (a) with $\mathcal{Y} = L_{\lambda_\diamond}$ and $T^* = \text{Id}$, taking advantage of Remarks 4.3.

$\qquad \square$

5. ENTROPIC PROJECTIONS

The results of the preceding sections are translated in terms of entropic projections. We consider the problem $(\mathcal{P}_{\mathcal{C}})$ where $\mathcal{C}$ is a convex subset of $M_{\mathcal{Z}}$.

5.1. Generalized entropic projections.

Let us start recalling an interesting result of Csiszár.

Definition 5.1. Let Q and $(Q_n)_{n \geq 1}$ in $M_{\mathcal{Z}}$ be absolutely continuous with respect to R.

(1) One says that $(Q_n)_{n \geq 1}$ converges in R-measure to Q if $\frac{\mathrm{d}Q_n}{\mathrm{d}R}$ converges in R-measure to $\frac{\mathrm{d}Q}{\mathrm{d}R}$.

(2) One says that $(Q_n)_{n \geq 1}$ converges $\sigma(L_{\lambda_\diamond^*}R, E_{\lambda_\diamond})$ to Q in $L_{\lambda_\diamond^*}R$ if $\frac{\mathrm{d}Q_n}{\mathrm{d}R}$ $\sigma(L_{\lambda_\diamond^*}R, E_{\lambda_\diamond})$-converges to $\frac{\mathrm{d}Q}{\mathrm{d}R}$.

(3) One says that $(Q_n)_{n \geq 1}$ converges in variation to Q if $\frac{\mathrm{d}Q_n}{\mathrm{d}R}$ converges to $\frac{\mathrm{d}Q}{\mathrm{d}R}$ in L_1.

Definition 5.2 (generalized entropic projection [5], Csiszár)**.** Suppose that $\mathcal{C} \cap \text{dom}\,I \neq \emptyset$ and that any minimizing sequence of the problem $(\mathcal{P}_{\mathcal{C}})$ converges in variation to some $Q_* \in M_{\mathcal{Z}}$. This Q_* is called the *generalized I-projection* of mR on $\mathcal{C}$ with respect to I. It may not belong to $\mathcal{C}$. In case Q_* is in $\mathcal{C}$, it is called the *I-projection* of mR on $\mathcal{C}$.

Theorem 5.3 (Csiszár [6])**.** *Suppose that* (A_R) *and* (A_{γ^*}) *are satisfied. Then, mR has a generalized I-projection on any convex subset $\mathcal{C}$ of $M_{\mathcal{Z}}$ such that $\mathcal{C} \cap \text{dom}\,I \neq \emptyset$.*

In [6] γ^* doesn't depend on the variable z, but the proof remains unchanged with a z-dependence.

Csiszár's proof of Theorem 5.3 is based on a parallelogram identity which allows to show that any minimizing sequence is a Cauchy sequence. This result is general but it doesn't tell much about the nature of Q_*. Let us give some details on the generalized entropic projections in specific situations.

As a direct consequence of Theorem 2.7, if the constraints are good, the generalized entropic projection is the entropic projection.

Proposition 5.4. *Suppose that* (A), (A$_\theta^\forall$) *and* (4.4) *hold, C is convex, $\mathcal{C}$ given at* (4.3) *is $\sigma(L_{\lambda_\diamond^*}R, E_{\lambda_\diamond})$-closed and $\mathcal{C} \cap \operatorname{dom} I \neq \emptyset$. Then, the I-projection Q_* exists and is equal to*

$$Q_* = \widehat{Q} \in \mathcal{C}$$

where $\widehat{Q}$ is the minimizer of (P$_C$) *which is described at Theorem* 2.7.
  Moreover, any minimizing sequence $\sigma(L_{\lambda_\diamond^*}R, L_{\lambda_\diamond})$-converges to Q_*.

Proof. This is an easy corollary of Theorem 2.7.   □

As a direct corollary of Proposition 5.4 and Remarks 4.3, we obtain the following result which is essentially [6] (Thm. 3(iii)).

Corollary 5.5. *Suppose that* (A) *and* (4.4) *hold. Let $\mathcal{C}$ be any $\sigma(L_{\lambda_\diamond^*}R, E_{\lambda_\diamond})$-closed convex set such that $\mathcal{C} \cap \operatorname{dom} I \neq \emptyset$. Then, the I-projection of mR on $\mathcal{C}$ exists.*

Theorem 5.6. *Suppose that* (A) *and* (4.4) *hold, $\mathcal{C}$ given at* (4.3) *is $\sigma(L_{\lambda_\diamond^*}R, L_{\lambda_\diamond})$-closed and $\mathcal{C} \cap \operatorname{dom} I \neq \emptyset$. Let us consider the additional conditions:*

  (1) (a) *There are finitely many moment constraints, i.e. $\mathcal{X}_o = \mathbb{R}^K$ (see Sect.* 3.2*)*
      (b) *$\mathcal{C} \cap \operatorname{icordom} I \neq \emptyset$;*
  (2) *$\mathcal{C}$ is a $\sigma(L_{\lambda_\diamond^*}R, L_{\lambda_\diamond})$-closed convex set with a nonempty $\|\cdot\|_{\lambda_\diamond^*}$-interior.*

Then, under one of the conditions (1) or (2), the generalized I-projection Q_* of mR on $\mathcal{C}$ is

$$Q_* = Q_\diamond$$

the absolutely continuous component described at (4.1) *and $I(Q_*) \leq \inf_\mathcal{C} I$.*

Proof. This is a direct consequence of Theorem 4.4.   □

5.2. **The special case of relative entropy**

The relative entropy $I(P|R)$ and its extension $\bar{I}(\ell|R)$ are described at Section 3. The minimization problem is

$$\text{minimize } I(P|R) \text{ subject to } \int_\mathcal{Z} \theta \, \mathrm{d}P \in C, \quad P \in P_\mathcal{Z} \tag{5.1}$$

and its extension is

$$\text{minimize } \bar{I}(\ell|R) \text{ subject to } \langle \theta, \ell \rangle \in C, \quad \ell \in O_{\exp}. \tag{5.2}$$

Recall that $O_{\exp}$ is defined at (3.4). We introduce the Cramér transform of the image law of R by θ on $\mathcal{X}_o$:

$$\Xi(x) = \sup_{y \in \mathcal{Y}_o} \left\{ \langle y, x \rangle - \log \int_\mathcal{Z} \mathrm{e}^{\langle y, \theta \rangle} \, \mathrm{d}R \right\} \in [0, \infty], \quad x \in \mathcal{X}_o. \tag{5.3}$$

Proposition 5.7 (relative entropy subject to good constraints)**.** *Let us assume that θ satisfies the "good constraint" assumption*

$$\forall y \in \mathcal{Y}_o, \int_\mathcal{Z} \mathrm{e}^{\langle y, \theta(z) \rangle} \, R(\mathrm{d}z) < \infty \tag{5.4}$$

and that $C \cap \mathcal{X}$ is a $\sigma(\mathcal{X}, \mathcal{Y})$-closed convex subset of $\mathcal{X}$.
(a) The following dual equality holds:

$$\inf\{I(P|R); \langle \theta, P \rangle \in C, P \in P_\mathcal{Z}\} = \sup_{y \in \mathcal{Y}_o} \left\{ \inf_{x \in C} \langle y, x \rangle - \log \int_\mathcal{Z} \mathrm{e}^{\langle y, \theta \rangle} \, \mathrm{d}R \right\} \in [0, \infty].$$

(b) *Suppose that in addition $C \cap \operatorname{dom} \Xi \neq \emptyset$. Then, the minimization problem* (5.1) *has a unique solution $\widehat{P}$ in $P_{\mathcal{Z}}$, $\widehat{P}$ is the entropic projection of R on $\mathcal{C} = \{P \in P_{\mathcal{Z}}, \int_{\mathcal{Z}} \theta \, dP \in C\}$.*

(c) *Suppose that in addition, $C \cap \operatorname{icordom} \Xi \neq \emptyset$, then there exists some linear form $\tilde{\omega}$ on $\mathcal{X}_o$ such that $\langle \tilde{\omega}, \theta \rangle$ is measurable and*

$$\begin{cases} \hat{x} := \int_{\mathcal{Z}} \theta \, d\widehat{P} \in C \cap \operatorname{dom} \Xi \\ \langle \tilde{\omega}, \hat{x} \rangle \leq \langle \tilde{\omega}, x \rangle, \forall x \in C \cap \operatorname{dom} \Xi \\ \widehat{P}(dz) = \exp\left(\langle \tilde{\omega}, \theta(z) \rangle - \log \int_{\mathcal{Z}} e^{\langle \tilde{\omega}, \theta \rangle} \, dR \right) \, R(dz). \end{cases} \tag{5.5}$$

In this situation, $\hat{x}$ minimizes Ξ on C, $I(\widehat{P} \mid R) = \Xi(\hat{x})$ and

$$\hat{x} = \int_{\mathcal{Z}} \theta(z) \exp\left(\langle \tilde{\omega}, \theta(z) \rangle - \log \int_{\mathcal{Z}} e^{\langle \tilde{\omega}, \theta \rangle} \, dR \right) \, R(dz) \tag{5.6}$$

in the weak sense. Moreover, $\tilde{\omega}$ satisfies (5.5) *if and only if it is the solution to*

$$maximize \quad \inf_{x \in C \cap \mathcal{X}} \langle \omega, x \rangle - \log \int_{\mathcal{Z}} e^{\langle \omega, \theta \rangle} \, dR, \quad \omega \in \widetilde{\mathcal{Y}} \tag{5.7}$$

where $\widetilde{\mathcal{Y}}$ is the cone of all measurable functions u such that u_- is in the $\| \cdot \|_1$-closure of $\langle \mathcal{Y}_o, \theta \rangle_-$ and u_+ is in the $\sigma(L_\tau, L_{\tau^*})$-closure of $\langle \mathcal{Y}_o, \theta \rangle_+$.

Proposition 5.8 (relative entropy subject to critical constraints)**.** *Let us assume that θ satisfies the "critical constraint" assumption*

$$\forall y \in \mathcal{Y}_o, \exists \alpha > 0, \int_{\mathcal{Z}} e^{\alpha |\langle y, \theta(z) \rangle|} \, R(dz) < \infty \tag{5.8}$$

and that $C \cap \mathcal{X}$ is a $\sigma(\mathcal{X}, \mathcal{Y})$-closed convex subset of $\mathcal{X}$.

(a) *The following dual equality holds:*

$$\inf\{\bar{I}(\ell \mid R); \langle \theta, \ell \rangle \in C, \ell \in O_{\exp}\} = \sup_{y \in \mathcal{Y}_o} \left\{ \inf_{x \in C} \langle y, x \rangle - \log \int_{\mathcal{Z}} e^{\langle y, \theta \rangle} \, dR \right\} \in [0, \infty].$$

(b) *Suppose that in addition $C \cap \operatorname{dom} \Xi \neq \emptyset$. Then, the minimization problem* (5.2) *is attained in $O_{\exp}$: the set of minimizers is nonempty, convex and $\sigma(L'_\tau, L_\tau)$-compact. Moreover, all the minimizers share the same unique absolutely continuous part $P_\diamond \in P_{\mathcal{Z}} \cap L_{\tau^*} R$ which is the generalized entropic projection of R on $\mathcal{C}$.*

(c) *Suppose that in addition, one of the following conditions*

    (1) $\mathcal{X}_o = \mathbb{R}^K$ and $C \cap \operatorname{icordom} I \neq \emptyset$ or

    (2) $\| \cdot \|_{\tau^*}\text{-int}(\mathcal{C}) \neq \emptyset$.

    is satisfied. Then, there exists a linear form $\tilde{\omega}$ on $\mathcal{X}_o$ such that $\langle \tilde{\omega}, \theta \rangle$ is measurable, $\int_{\mathcal{Z}} e^{\langle \tilde{\omega}, \theta \rangle} \, dR < \infty$ and

$$P_\diamond(dz) = \exp\left(\langle \tilde{\omega}, \theta(z) \rangle - \log \int_{\mathcal{Z}} e^{\langle \tilde{\omega}, \theta \rangle} \, dR \right) \, R(dz). \tag{5.9}$$

    Moreover, $\tilde{\omega}$ is the solution to (5.7)*.*

In Proposition 5.7, $\hat{x}$ is the dominating point in the sense of Ney (see Def. 6.2) of C with respect to Ξ. The representation of $\hat{x}$ has already been obtained for C with a nonempty topological interior by Ney [23] and Einmahl and Kuelbs [10], respectively in $\mathbb{R}^d$ and in a Banach space setting. The representation of the generalized projection $P_\diamond$ is obtained with a very different proof by Csiszár [4] and ([5], Thm. 3). Proposition 5.8 also extends corresponding results of Kuelbs [14] which are obtained in a Banach space setting.

Proof of Propositions 5.7 and 5.8. They are direct consequences of Theorems 2.7, 2.8, Proposition 5.4, Theorem 5.6 and Lemma 5.9 below. □

The following lemma allows to apply the results of the present paper with $\gamma(s) = \mathrm{e}^s - 1$ and the extended constraint $\langle(\mathbf{1}, \theta), \ell\rangle \in \{1\} \times C$: the first component of the constraint insures the unit mass $\langle\mathbf{1}, \ell\rangle = 1$, to obtain results in terms of log-Laplace transform.

Lemma 5.9. *For all $x \in \mathcal{X}_o$,*

$$\sup_{y \in \mathcal{Y}_o}\left\{\langle y, x\rangle - \log\int_{\mathcal{Z}}\mathrm{e}^{\langle y, \theta\rangle}\,\mathrm{d}R\right\} = \sup_{\tilde{y} \in \mathbb{R} \times \mathcal{Y}_o}\left\{\langle\tilde{y}, (1, x)\rangle - \int_{\mathcal{Z}}(\mathrm{e}^{\langle\tilde{y}, (1, \theta)\rangle} - 1)\,\mathrm{d}R\right\} \in (-\infty, +\infty].$$

Proof. Using the identity: $-\log b = \sup_a\{a + 1 - b\mathrm{e}^a\}$, we get:

$$
\begin{aligned}
\sup_{y \in \mathcal{Y}_o}\left\{\langle y, x\rangle - \log\int_{\mathcal{Z}}\mathrm{e}^{\langle y, \theta\rangle}\,\mathrm{d}R\right\} &= \sup_{a \in \mathbb{R}, y \in \mathcal{Y}_o}\left\{\langle y, x\rangle + a + 1 - \mathrm{e}^a\int_{\mathcal{Z}}\mathrm{e}^{\langle y, \theta\rangle}\,\mathrm{d}R\right\} \\
&= \sup_{a \in \mathbb{R}, y \in \mathcal{Y}_o}\left\{\langle(a, y), (1, x)\rangle - \int_{\mathcal{Z}}\mathrm{e}^{\langle y, \theta\rangle + a}\,\mathrm{d}R + 1\right\} \\
&= \sup_{\tilde{y} \in \mathbb{R} \times \mathcal{Y}_o}\left\{\langle\tilde{y}, (1, x)\rangle - \int_{\mathcal{Z}}(\mathrm{e}^{\langle\tilde{y}, (1, \theta)\rangle} - 1)\,\mathrm{d}R\right\}
\end{aligned}
$$

which is the desired result. $\square$

Formula (5.3) states that $\Xi = \Lambda^*$ where

$$\Lambda(y) = \log\int_{\mathcal{Z}}\mathrm{e}^{\langle y, \theta\rangle}\,\mathrm{d}R, \quad y \in \mathcal{Y}_o \tag{5.10}$$

is the log-Laplace transform of the image of R by the mapping θ. On the other hand, Lemma 5.9 states that $\Lambda^*(x) = \Gamma^*(1, x)$. Hence, when working with the relative entropy, one can switch from Γ to Λ.

Example 5.10. *Csiszár's example.* Comparing Proposition 5.4 with Theorem 5.6, one may wonder if the $\sigma(L_{\lambda_\diamond^*}R, E_{\lambda_\diamond})$-closedness of the convex set $\mathcal{C}$ is critical for the existence of an entropic projection. The answer is affirmative. In [5] (ex. 3.2), Csiszár gives an interesting example where the generalized entropic projection can be explicitly computed in a situation where (P_C) is not attained, see also [9] (exercise 7.3.11). This example is the following one.

Take the probability measure on $\mathcal{Z} = [0, \infty)$ defined by $R(\mathrm{d}z) = \frac{1}{a_0}\frac{\mathrm{e}^{-z}}{1+z^3}\,\mathrm{d}z$ where a_0 is the normalizing constant, I the relative entropy with respect to R, $\theta : z \in [0, \infty) \mapsto z \in \mathcal{X} = \mathbb{R}$ and $C = [c, \infty)$. This gives $\mathcal{C} = \{Q \in L_{\tau^*}R; \int_{[0, \infty)} z\,Q(\mathrm{d}z) \geq c, Q([0, \infty)) = 1\}$.

From the point of view of the present article, the main point is that θ is in $L_\tau(R)$ but not in $E_\tau(R)$. Indeed, the log-Laplace transform is

$$\Lambda(y) = \log\int_{[0, \infty)}\mathrm{e}^{(y-1)z}\frac{\mathrm{d}z}{a_0(1+z^3)}, \quad y \in \mathbb{R} \tag{5.11}$$

whose effective domain is $(-\infty, 1]$ which admits 0 as an interior point but is not the whole line. A graphic representation of Λ is drawn at Figure 3 below, see example 6.9.

By Theorem 5.8, the generalized projection of R on $\mathcal{C}$ exists and is equal to $P_y = \frac{1}{a_y}\frac{\mathrm{e}^{(y-1)z}}{1+z^3}\,\mathrm{d}z$ for some real $y \leq 1$, where $a_y = a_0\mathrm{e}^{\Lambda(y)}$ is the normalizing constant. If it belongs to $\mathcal{C}$, then $\Lambda'(y) = \int_{\mathcal{Z}} z\,P_y(\mathrm{d}z) \geq c$. But Λ is convex so that Λ' is increasing and $\sup_{y \leq 1}\int_{\mathcal{Z}} z\,P_y(\mathrm{d}z) = \int_{\mathcal{Z}} z\,P_1(\mathrm{d}z) = \frac{1}{a_1}\int_{\mathcal{Z}}\frac{z}{1+z^3}\,\mathrm{d}z := x_* < \infty$. Therefore, for any $c > x_*$, there are no entropic projection but only a generalized one which is $P_1(\mathrm{d}z) = \frac{\mathrm{d}z}{a_1(1+z^3)}$.

A detailed analysis of this example in terms of singular component is done by Léonard and Najim [20] (Prop. 3.9). This example corresponds to a $\sigma(L_{\lambda_\diamond^*}R, L_{\lambda_\diamond})$-closed convex set $\mathcal{C}$ such that no entropic projection exists but only a generalized one.

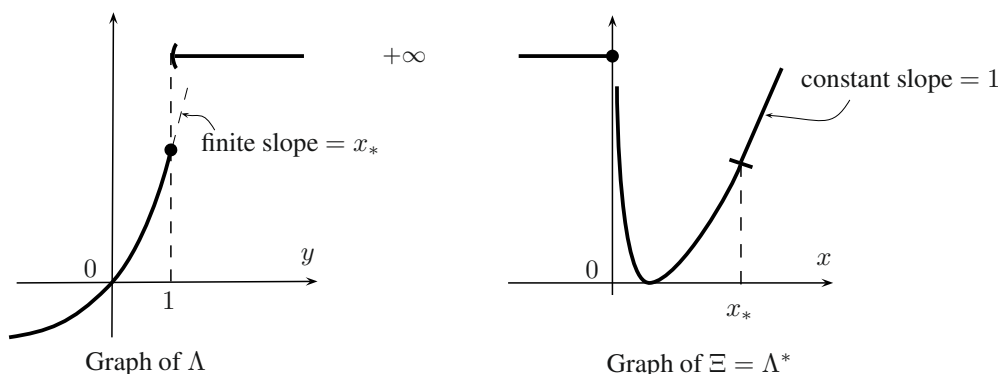More details about this example are given below at examples 6.9 and 7.2.

FIGURE 3. The case: $R(\mathrm{d}z) = \frac{1}{a_0} \frac{\mathrm{e}^{-z}}{1+z^3} \, \mathrm{d}z$.

6. DOMINATING POINTS

Following Ney [22,23], let us introduce the following definition. A point $\hat{x} \in \mathcal{X}_o$ sharing the properties (2.15)(a,b) and (2.16) of Theorem 2.7 is called a *dominating point*.

Definition 6.1 (dominating point)**.** Let $C \subset \mathcal{X}_o$ be a convex set such that $C \cap \mathcal{X}$ is $\sigma(\mathcal{X}, \mathcal{Y})$-closed. The point $\hat{x} \in \mathcal{X}_o$ is called a Γ^*-*dominating point* if

(a) $\hat{x} \in C \cap \operatorname{dom} \Gamma^*$,
(b) there exists some linear form $\tilde{\omega}$ on $\mathcal{X}_o$ such that $\langle \tilde{\omega}, \hat{x} \rangle \leq \langle \tilde{\omega}, x \rangle$ for all $x \in C \cap \operatorname{dom} \Gamma^*$,
(c) $\langle \tilde{\omega}, \theta(\cdot) \rangle$ is measurable and $\hat{x} = \int_{\mathcal{Z}} \theta(z) \gamma'(z, \langle \tilde{\omega}, \theta(z) \rangle) \, R(\mathrm{d}z)$.

In the special case where Γ^* is replaced by the Cramér transform Ξ defined at (5.3), taking Lemma 5.9 into account, this definition becomes the following one.

Definition 6.2. Let $C \subset \mathcal{X}_o$ be a convex set such that $C \cap \mathcal{X}$ is $\sigma(\mathcal{X}, \mathcal{Y})$-closed. The point $\hat{x} \in \mathcal{X}_o$ is called a Ξ-*dominating point* if

(a) $\hat{x} \in C \cap \operatorname{dom} \Xi$;
(b) there exists some linear form $\tilde{\omega}$ on $\mathcal{X}_o$ such that $\langle \tilde{\omega}, \hat{x} \rangle \leq \langle \tilde{\omega}, x \rangle$, for all $x \in C \cap \operatorname{dom} \Xi$ and
(c) $\langle \tilde{\omega}, \theta(\cdot) \rangle$ is measurable and $\hat{x} = \int_{\mathcal{Z}} \theta(z) \dfrac{\exp(\langle \tilde{\omega}, \theta(z) \rangle)}{Z(\tilde{\omega})} \, R(\mathrm{d}z)$ where $Z(\tilde{\omega})$ is the unit mass normalizing constant.

Note that this definition is slightly different from the ones proposed by Ney [23] and Einmahl and Kuelbs [10] since C is neither supposed to be an open set nor to have a non-empty interior and $\hat{x}$ is not assumed to be a boundary point of C. The above integral representation (c) is (5.6).

We are going to investigate some relations between dominating points and entropic projections. In the case where the constraint is good, Theorem 2.7 and Proposition 5.4 state that the generalized entropic projection is the entropic projection $Q_* = \widehat{Q}$, the minimizer $\bar{x}$ is the dominating point of C, it is related to Q_* by the identity:

$$\bar{x} = \langle \theta, Q_* \rangle. \tag{6.1}$$

We now look at the situation where the constraint is critical. As remarked at example 5.10, the above equality may fail. A necessary and sufficient condition (in terms of the function Γ^*) for $\bar{x}$ to satisfy (6.1) is obtained at Theorem 6.7.

Recall that the extended entropy $\bar{I}$ is given by (2.9): $\bar{I}(\ell) = I(\ell^a) + I^s(\ell^s)$, $\ell = \ell^a + \ell^s \in L'_{\lambda_\diamond}$ and define for all $x \in \mathcal{X}_o$,

$$
\begin{aligned}
\bar{J}(x) &:= \inf\{\bar{I}(\ell); \ell \in L'_{\lambda_\diamond}, \langle \theta, \ell \rangle = x\} \\
J(x) &:= \inf\{I(\ell); \ell \in L_{\lambda_\diamond^*} R, \langle \theta, \ell \rangle = x\} \\
J^s(x) &:= \inf\{I^s(\ell); \ell \in L^s_{\lambda_\diamond}, \langle \theta, \ell \rangle = x\}.
\end{aligned}
$$

Because of the decomposition $L'_{\lambda_\diamond} \simeq L_{\lambda_\diamond^*} R \oplus L^s_{\lambda_\diamond}$, we obtain for all $x \in \mathcal{X}_o$,

$$
\begin{aligned}
\bar{J}(x) &= \inf\{I(\ell_1) + I^s(\ell_2); \ell_1 \in L_{\lambda_\diamond^*} R, \ell_2 \in L^s_{\lambda_\diamond}, \langle \theta, \ell_1 + \ell_2 \rangle = x\} \\
&= \inf\{J(x_1) + J^s(x_2); x_1, x_2 \in \mathcal{X}_o, x_1 + x_2 = x\} \\
&= J \square J^s(x)
\end{aligned}
$$

where $J \square J^s$ is the inf-convolution of J and J^s. By Theorem 2.8a, $\bar{J} = \Gamma^*$ and if $\bar{J}(x) < \infty$, there exists $\ell_x \in L'_{\lambda_\diamond}$ such that $\langle \theta, \ell_x \rangle = x$ and $\bar{J}(x) = \bar{I}(\ell_x)$. Let us define

$$
x^a := \langle \theta, \ell_x^a \rangle \quad \text{and} \quad x^s := \langle \theta, \ell_x^s \rangle.
$$

These definitions make sense since ℓ_x^a is the unique (common) absolutely continuous part of the minimizers of $\bar{I}$ on the closed convex set $\{\ell \in L'_{\lambda_\diamond}; \langle \theta, \ell \rangle = x\}$, see Theorem 4.1a. Of course, we have

$$
x = x^a + x^s
$$

and as $\bar{J}(x) = \bar{I}(\ell_x) = I(\ell_x^a) + I^s(\ell_x^s) \geq J(x^a) + J^s(x^s) \geq J \square J^s(x) = \bar{J}(x)$, we get the following result.

Proposition 6.3. *For all $x \in \operatorname{dom} \bar{J}$, we have:*

$$
\bar{J}(x) = J(x^a) + J^s(x^s), J(x^a) = I(\ell_x^a) \text{ and } J^s(x^s) = I^s(\ell_x^s).
$$

Now, let us have a look at the dual equalities. The *recession function* of Γ^* is defined for all x by

$$
\widetilde{\Gamma^*}(x) := \lim_{t \to +\infty} \Gamma^*(tx)/t \in (-\infty, +\infty]. \tag{6.2}
$$

Definition 6.4 (recessive x)**.** Let us say that x is *recessive* for Γ^* if for some $\delta > 0$ and $\xi \in \mathcal{X}_o$, $\Gamma^*(x + t\xi) - \Gamma^*(x) = t\widetilde{\Gamma^*}(\xi)$ for all $t \in (-\delta, +\infty)$. It is said to be *non-recessive* otherwise.

Proposition 6.5. *We have $\bar{J} = \Gamma^*$ and $J^s = \widetilde{\Gamma^*}$. Moreover, $J(x) = \Gamma^*(x)$ for all non-recessive $x \in \mathcal{X}_o$.*

Proof. We have already noted that $\bar{J} = \Gamma^*$ and by [17] (Thm. 2.3), we get: $J^s = \iota^*_{\operatorname{dom} \Gamma}$: the support function of $\operatorname{dom} \Gamma$. Therefore, it is also the recession function of Γ^*. Hence, we have $\Gamma^* = \bar{J} = J \square J^s = J \square \widetilde{\Gamma^*}$.

Comparing $\Gamma^* = J \square \widetilde{\Gamma^*}$ with the general identity $\Gamma^* = \Gamma^* \square \widetilde{\Gamma^*}$, we obtain that $J(x) = \Gamma^*(x)$, for all non-recessive $x \in \mathcal{X}_o$. $\square$

Proposition 6.6. *For all $x \in \operatorname{dom} \Gamma^*$, we have:*

$$
\Gamma^*(x) = \Gamma^*(x^a) + \widetilde{\Gamma^*}(x^s).
$$

Moreover, x is non-recessive if and only if $x^s = 0$. In particular, x^a is non-recessive.

Proof. By (2.9), we have $\bar{I}(\ell_x) = I(\ell_x^a) + \widetilde{I}(\ell_x^s)$ where $\widetilde{I}$ is the recession function of I. It follows that $\bar{J}(x) = J(x_a) + J^s(x^s)$, since $J(x_a) = I(\ell_x^a)$ (Prop. 6.3) and the recession function of $\bar{J}$ is J^s. To show this, note that (see [27])

    – I^s is the recession function of $\bar{I}$,
    – the epigraph of $x \mapsto \inf\{f(\ell); \ell, T\ell = x\}$ (with T a linear operator) is "essentially" a linear projection of the epigraph of f, (let us call it an inf-projection),
    – the epigraph of the recession function is the recession cone of the epigraph and
    – the inf-projection of a recession cone is the recession cone of the inf-projection.

The first result now follows from $\bar{J} = \Gamma^*$. The same set of arguments also yields the second statement.   □

Theorem 6.7. *Let us assume that the hypotheses of Theorem 2.8 hold and that $C \cap \operatorname{icordom}\Gamma^* \neq \emptyset$.*

(a) *Then, a minimizer $\bar{x}$ of Γ^* on the set C is a Γ^*-dominating point of C if and only if $\bar{x}$ is non-recessive. This is also equivalent to the following statement: "all the solutions of the minimization problem $(\overline{\mathrm{P}}_C)$ are absolutely continuous with respect to R." In such a case the solution of $(\overline{\mathrm{P}}_C)$ is unique and it matches the solution of (P_C).*

(b) *In particular when Γ^* admits a degenerate recession function, i.e. $\widetilde{\Gamma^*}(x) = +\infty$ for all $x \neq 0$, then the minimizer $\bar{x}$ is a Γ^*-dominating point of C.*

(c) *The same statements hold with Γ^* replaced by Ξ.*

Proof. This is a direct consequence of Theorem 2.8, Proposition 6.6 and Lemma 5.9.   □

Remark 6.8. *A remark about the steepness of the log-Laplace transform.* In [14] (Thm. 1), with the setting of Section 5.2 where $\mathcal{X}_o$ is a Banach space, Kuelbs proves a result that is slightly different from statement (b) of the above theorem. It is proved that the existence of a Ξ-dominating point for *all* convex sets C with a nonempty topological interior is equivalent to some property of the Gâteaux derivative of the log-Laplace transform $y \in \mathcal{X}_o' \mapsto \Lambda(y) = \log \int_{\mathcal{X}_o} \exp(\langle y, x \rangle) R \circ \theta^{-1}(\mathrm{d}x)$ on the boundary of its domain. This property is an infinite dimensional analogue of the steepness of the log-Laplace transform. It turns out that it is equivalent to the following assumption: the Cramér transform $\Xi = \Lambda^*$ admits a degenerate recession function.

Example 6.9. *Csiszár's example continued.* Recall that by Lemma 5.9, $\Xi(x) = \Lambda^*(x) = \Gamma^*(1, x)$ so that Γ and Λ play the same role. Clearly, $\operatorname{dom}\Lambda = (-\infty, 1]$ and $\Lambda'(1^-) = \int_{[0,\infty)} z\, P_1(\mathrm{d}z) := x_* < \infty$. It follows that for all $x \geq x_*$, $\Xi(x) - \Xi(x_*) = x - x_*$.

We deduce from this that (x_*, ∞) is a set of recessive points. By Theorem 6.7, they cannot be dominating points. Note also that the log-Laplace transform Λ is not steep.

Remark 6.10. *A remark about Csiszár's example.* For comparison, suppose that R in Csiszár's example 5.10 and 6.9 is replaced by its slight modification $R(\mathrm{d}z) = \mathrm{e}^{-z}\,\mathrm{d}z$: the exponential law with parameter 1. Then $\Lambda(y) = -\log(1-y) + \iota_{\{y \geq 1\}}$ and $\Xi(x) = \Lambda^*(x) = x - 1 - \log x + \iota_{\{x \leq 0\}}$. Their graphic representations are given at Figure 4 and one sees that for any $c > 0$, R admits the entropic projection $P_y(\mathrm{d}z) = (1-y)\mathrm{e}^{(y-1)z}\,\mathrm{d}z$ which is the exponential law with parameter $(1-y)$ with y such that $1/(1-y) = \int_{[0,\infty)} z\, P_y(\mathrm{d}z) = c$. In this case, for any $c > 0$ the generalized entropic projection is the entropic projection, Λ is steep and Ξ has no recessive points. Notice however that the constraint is critical as in example 6.9.

7. CONDITIONAL LAWS OF LARGE NUMBERS

In this last section, we give a probabilistic interpretation of the singular component of the generalized I-projection in terms of a conditional LLN.
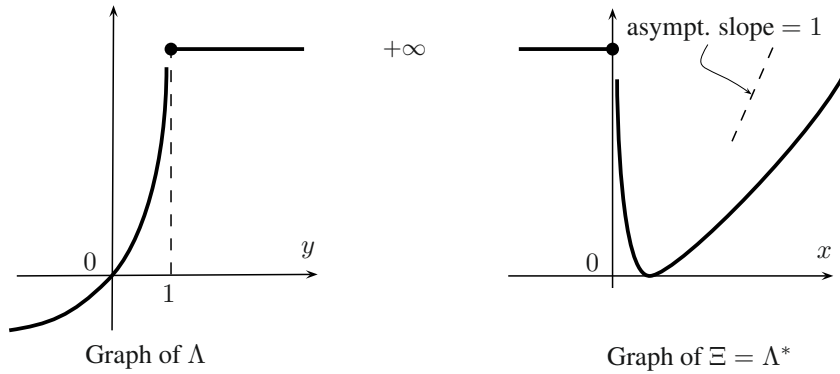
Graph of Λ Graph of $\Xi = \Lambda^*$

FIGURE 4. The case: $R(\mathrm{d}z) = \mathrm{e}^{-z}\,\mathrm{d}z$.

7.1. Conditional laws of large numbers

Conditional laws of large numbers are already well-known, see [1] or [30] for instance. Theorem 7.1 below is a general statement of a conditional LLN which is inspired from [30] and [9] (Sect. 7.3). Its proof is given at Appendix B.

Let $\{L_n\}$ be a sequence of random vectors in the algebraic dual space $\mathcal{L}_o$ of some vector space $\mathcal{U}_o$. As a typical instance, one can think of random measures L_n with $\mathcal{U}_o$ a function space. Let $T : \mathcal{L}_o \to \mathcal{X}_o$ be a linear operator with values in another vector space $\mathcal{X}_o$. We are going to investigate the behavior of the conditional law $\mathbb{P}(L_n \in \cdot \mid TL_n \in C)$ of L_n as n tends to infinity, for some measurable set C in $\mathcal{X}_o$. It appears that this type of conditional law of large numbers is connected with large deviations. We assume that $\{L_n\}$ obeys the LDP with a good rate function I in $\mathcal{L}_o$ endowed with the weak topology $\sigma(\mathcal{L}_o, \mathcal{U}_o)$ and the associated cylinder σ-field. It is also clear that one should assume that $\mathbb{P}(TL_n \in C) > 0$ for all n, not to divide by zero. To overcome this restriction, we look at $\mathbb{P}(L_n \in \cdot \mid TL_n \in C_\delta)$ where C_δ tends to C as δ tends to zero.

Let us assume that $\mathcal{X}_o$ is a topological vector space with its Borel σ-field and that $T : \mathcal{L}_o \to \mathcal{X}_o$ is continuous. The contraction principle tells us that

$$X_n := TL_n$$

obeys the LDP in $\mathcal{X}_o$ with the rate function

$$J(x) = \inf\{I(\ell); \ell \in \mathcal{L}_o, T\ell = x\}.$$

The following set of hypotheses is based on a framework which appears in [30] and [9] (Sect. 7.3).

7.2. Assumptions (B).

(B_L) The sequence $\{L_n\}$ obeys the LDP in $\mathcal{L}_o$ with a *good* rate function I. This means that I is inf-compact.
(B_T) The linear operator $T : \mathcal{L}_o \to \mathcal{X}_o$ is continuous.
(B_C) *Assumptions on C.* As a convention, we write $J(C)$ for $\inf_{x \in C} J(x)$.
    (1) $J(C) < \infty$.
    (2) The set C is closed. It is the limit as δ decreases to 0: $C = \cap_\delta C_\delta$, of a decreasing family of closed sets C_δ in $\mathcal{X}_o$ such that for all $\delta > 0$ and all $n \geq 1$, $\mathbb{P}(X_n \in C_\delta) > 0$
    (3) and one of the following statements
        (a) $C_\delta = C$ for all $\delta > 0$ and $J(\mathrm{int}\, C) = J(C)$ or
        (b) $C \subset \mathrm{int}\, C_\delta$ for all $\delta > 0$
    is fulfilled.

Let $\mathcal{G}$ be the set of all solutions of the following minimization problem:

$$\text{minimize } I(\ell) \text{ subject to } T\ell \in C, \quad \ell \in \mathcal{L}_o. \tag{7.1}$$

Similarly, let $\mathcal{H}$ be the set of all solutions of the following minimization problem:

$$\text{minimize } J(x) \text{ subject to } x \in C, \quad x \in \mathcal{X}_o. \tag{7.2}$$

We can now state a result about conditional laws of large numbers which is proved at the Appendix.

Theorem 7.1. *Assume* (B). *For all open subset G of $\mathcal{L}_o$ such that $\mathcal{G} \subset G$ and all open subset H of $\mathcal{X}_o$ such that $\mathcal{H} \subset H$, we have*

$$\limsup_{\delta \to 0} \limsup_{n \to \infty} \frac{1}{n} \log \mathbb{P}(L_n \notin G \mid TL_n \in C_\delta) < 0 \text{ and}$$

$$\limsup_{\delta \to 0} \limsup_{n \to \infty} \frac{1}{n} \log \mathbb{P}(X_n \notin H \mid X_n \in C_\delta) < 0.$$

In particular, if C is convex and the rate functions I and J are strictly convex, we have the conditional laws of large numbers:

$$\lim_{\delta} \lim_{n} \mathbb{P}(L_n \in \cdot \mid TL_n \in C_\delta) = \delta_{\bar{\ell}}$$

$$\lim_{\delta} \lim_{n} \mathbb{P}(X_n \in \cdot \mid X_n \in C_\delta) = \delta_{\bar{x}}$$

where the limits are understood with respect to the usual weak topologies of probability measures, $\bar{\ell}$ is the unique solution to the convex minimization problem (7.1) *and $\bar{x} = T\bar{\ell}$ is the unique solution to* (7.2).

7.3. **Empirical measures of independent samples**

Let $Z_1, Z_2, \ldots$ be an iid sequence of $\mathcal{Z}$-valued random variables which are R-distributed. The LDP for $L_n = \frac{1}{n} \sum_{i=1}^n \delta_{Z_i}$ is described by Sanov's theorem: L_n obeys the LDP in $P_{\mathcal{Z}}$ endowed with the topology $\sigma(P_{\mathcal{Z}}, B_{\mathcal{Z}})$ with the relative entropy $I(.|R)$ as its rate function, see [9] (Thm. 6.2.10). Here, $B_{\mathcal{Z}}$ denotes the space of all bounded measurable functions on $\mathcal{Z}$.

One can improve this result: this LDP still holds in $P_{\exp} \subset P_{\mathcal{Z}}$ with the topology $\sigma(P_{\exp}, \mathcal{E}_{\exp})$ where $\mathcal{E}_{\exp} = \{u; \forall a > 0, \int_{\mathcal{Z}} e^{a|u|} dR < \infty\}$ is $E_\tau(R)$ without identifying the R-a.e. equal functions and $P_{\exp} = \{P \in P_{\mathcal{Z}}; \int_{\mathcal{Z}} |u| dP < \infty, \forall u \in \mathcal{E}_{\exp}\}$, see [20,28].

A further improvement is obtained in [20]: L_n obeys the LDP in $O_{\exp}$ endowed with the topology $\sigma(O_{\exp}, \mathcal{L}_{\exp})$ with the extended relative entropy $\bar{I}(.|R)$ as its rate function. Here, $\mathcal{L}_{\exp} = \{u; \exists a > 0, \int_{\mathcal{Z}} e^{a|u|} dR < \infty\}$ is $L_\tau(R)$ without identifying the R-a.e. equal functions.

Hence, under the assumptions (B$_C$) on $\{C_\delta\}$ and C, we can apply Theorem 7.1 with $I = \bar{I}(.|R)$ and $J = \Xi$ which are defined at (3.3) and (5.3). More precisely, if the hypotheses of Proposition 5.7 are fulfilled, we see that

$$\lim_{\delta} \lim_{n} \mathbb{P}\left(\frac{1}{n} \sum_{i=1}^n \delta_{Z_i} \in \cdot \mid \frac{1}{n} \sum_{i=1}^n \theta(Z_i) \in C_\delta\right) = \delta_{\widehat{P}} \tag{7.3}$$

where $\widehat{P}$ is characterized by (5.5) and (5.7) and this limit corresponds to the usual weak topology on the set of probability measures on $P_{\exp}$ furnished with the topology $\sigma(P_{\exp}, \mathcal{E}_{\exp})$. If the hypotheses of Proposition 5.8 are fulfilled and one of the following assumptions

(1) $\mathcal{X}_o = \mathbb{R}^K$ and $\mathcal{C} \cap \operatorname{icordom} I \neq \emptyset$ or
(2) $\|\cdot\|_{\tau^*}\text{-int}(\mathcal{C}) \neq \emptyset$

holds, Theorem 7.1 and Proposition 5.8 tell us that (7.3) still holds, more precisely

$$\lim_{\delta} \lim_{n} \mathbb{P}\left(\frac{1}{n} \sum_{i=1}^n \delta_{Z_i} \in \cdot \mid \frac{1}{n} \sum_{i=1}^n \theta(Z_i) \in C_\delta\right) = \delta_{P_\diamond}$$

with respect to the same topology $\sigma(P_{\exp}, \mathcal{E}_{\exp})$, where $P_\diamond$ is characterized at (5.9). See [5] for similar results w.r.t. $\sigma(P_{\mathcal{Z}}, B_{\mathcal{Z}})$.

7.4. A probabilistic interpretation of the singular component

On the other hand, considering $L_n \in P_{\mathcal{Z}} \subset O_{\exp}$, (7.3) fails in $O_{\exp}$ endowed with topology $\sigma(O_{\exp}, \mathcal{L}_{\exp})$. Instead of this, the laws $\mathbb{P}\left(L_n \in \cdot \mid \frac{1}{n}\sum_{i=1}^n \theta(Z_i) \in C_\delta\right)$ admit cluster points which are probability measures on $O_{\exp}$ whose support consists of solutions $\hat{\ell}$ to the minimization problem (5.2). We know that $\hat{\ell}^a = P_\diamond$ for any $\hat{\ell}$.

By Proposition 2.3, $\langle \hat{\ell}^s, u \rangle = 0$ for all $u \in \mathcal{E}_{\exp}$. In particular, the mass of $\hat{\ell}^s$ is $\langle \hat{\ell}^s, \mathbf{1}\rangle = 0$. Nevertheless, $\hat{\ell}^s$ may be nonzero. This reflects the fact for some indices $1 \le i \le k_n$ (we choose the first ones without loss of generality) with a vanishing ratio: $\lim_{n\to\infty} k_n/n = 0$, and for some functions u in $\mathcal{L}_{\exp} \setminus \mathcal{E}_{\exp}$, $\frac{1}{n}\sum_{i=1}^{k_n} u(Z_i)$ does not vanish as n tends to infinity. Meanwhile, the remaining variables are such that $\lim_{n\to\infty} \frac{1}{n-k_n}\sum_{i=k_n+1}^{n} \delta_{Z_i} = P_\diamond$ with respect to $\sigma(O_{\exp}, \mathcal{L}_{\exp})$.

Example 7.2. Consider example 5.10 with $c > x_*$. Suppose that $L_n - \frac{1}{n}\delta_{Z_1}$ is close to $P_\diamond = P_1$ and $Z_1/n \ge c - x_*$ so that the event $\frac{1}{n}\sum_{i=1}^n Z_i \ge c - 1/n$ is realized. When n is large, this happens with a probability $p_n \simeq \exp[-(n-1)I(P_\diamond|R)]\mathbb{P}(Z_1 \ge n(c-x_*))$ and with the notation of example 6.9, $\lim_{n\to\infty} -\frac{1}{n}\log p_n = I(P_\diamond|R) + (c-x_*) = \Gamma^*(x_*) + [\Gamma^*(c) - \Gamma^*(x_*)] = \Gamma^*(c)$. Therefore, p_n has the optimal logarithmic behavior.

What is performed by the first particle $i = 1$, may also be performed by any collection of k_n different particles with a probability $p_n(a_1^n, \ldots, a_{k_n}^n) \simeq \binom{n}{k_n} \exp[-(n-k_n)I(P_\diamond|R)]\mathbb{P}(Z_1 \ge a_1^n) \cdots \mathbb{P}(Z_{k_n} \ge a_{k_n}^n)$ for any $a_1^n, \ldots, a_{k_n}^n$ such that

$$\begin{cases} \lim_{n\to\infty} k_n/n & = \quad 0, \\ \lim_{n\to\infty} a_k^n & = \quad +\infty, \quad \forall 1 \le k \le k_n, \\ \lim_{n\to\infty}(a_1^n + \cdots + a_{k_n}^n)/n & = \quad c - x_*. \end{cases}$$

Again, $p_n(a_1^n, \ldots, a_{k_n}^n)$ is logarithmically optimal since $\lim_{n\to\infty} -\frac{1}{n}\log p_n(a_1^n, \ldots, a_{k_n}^n) = I(P_\diamond|R) + \lim_{n\to\infty}(a_1^n + \cdots + a_{k_n}^n)/n = I(P_\diamond|R) + (c - x_*) = \Gamma^*(c)$.

A *vanishing ratio* of particles performing very large jumps is responsible for the appearance of the singular component.

For other examples of minimization problems (5.2) whose solutions exhibit a nonzero singular component, one can have a look at [19] (Sect. 7.4).

7.5. Empirical measures with random weights

In [8,12], one gives a Bayesian interpretation of the conditional LLN of the empirical measures with random weights $L_n = \frac{1}{n}\sum_{i=1}^n W_i \delta_{z_i}$, see (1.4). Let the weights W be iid random variables with a common Cramér transform γ^*. Let $\mathcal{Z}$ be a topological space with its Borel σ-field. Denote $C_{\mathcal{Z}}$ the space of all bounded continuous functions on $\mathcal{Z}$. The LDP for L_n is given at [9] (Thm. 7.2.3): if γ^* satisfies (4.4), L_n obeys the LDP in $M_{\mathcal{Z}}$ with the topology $\sigma(M_{\mathcal{Z}}, C_{\mathcal{Z}})$ and the rate function $I = I_{\gamma^*}$.

Under the hypotheses of Theorems 5.6 and 7.1, we obtain

$$\lim_\delta \lim_n \mathbb{P}\left(\frac{1}{n}\sum_{i=1}^n W_i \delta_{z_i} \in \cdot \mid \frac{1}{n}\sum_{i=1}^n W_i \theta(z_i) \in C_\delta\right) = \delta_{Q_\diamond}$$

where $Q_\diamond$ is described at (4.1) and the topology on $M_{\mathcal{Z}}$ is $\sigma(M_{\mathcal{Z}}, C_{\mathcal{Z}})$.

Under the more restrictive hypotheses of Proposition 5.4, we have $Q_\diamond = \widehat{Q}$ which is the solution of (P_C). If the weights W are nonnegative and $\mathbb{E}W = 1$, the reference probability measure R is interpreted as the prior distribution while $\widehat{Q}$ is a posterior distribution which is selected in $\mathcal{C}$ by means of (P_C). Hence, it is possible to estimate $\widehat{Q}$ with L_n. It is the aim of the method of Maximum Entropy in the Mean (MEM) to provide an effective simulation procedure of the posterior distribution.

The reverse entropy corresponds to a γ^* which doesn't satisfy (4.4); we have $L_{\lambda_\diamond} = L_\infty$ and $L'_{\lambda_\diamond} = L_1 R \oplus L'_\infty$. This situation is investigated in details in [12] with a compact space $\mathcal{Z}$. This compactness allows to restrict the attention to the subspace $M_{\mathcal{Z}} \subset L_1 R \oplus L'_\infty$ which is decomposed into the direct sum of the spaces of measures which are absolutely continuous and singular with respect to R. It is proved that the (δ, n)-sequence $\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^n W_i \delta_{z_i} \in \cdot \mid \frac{1}{n}\sum_{i=1}^n W_i \theta(z_i) \in C_\delta\right)$ admits cluster points in $M_{\mathcal{Z}}$ whose support consist of solutions of $(\overline{P}_C)$ which may have a singular component. The probabilistic interpretation of the Dirac part of this singular component is: a *non-vanishing ratio* of indices i accumulate in areas where the z_i are close to each other, with large values of W_i.

One can expect that L_n obeys LDPs with respect to topologies weakened by unbounded functions, but no results of this kind appear in the literature. If this holds true, for any γ^* satisfying (4.4) but such that $\lambda_\diamond$ isn't Δ_2-regular, an interpretation of the singular component in terms a vanishing ratio of indices is still available, as in Section 7.4.

APPENDIX A. COMPLEMENTS FOR THE STATEMENTS OF THEOREMS 2.7 AND 2.8

The decomposition into positive and negative parts of linear forms is necessary to state the extended dual problem which is needed for the characterization of the minimizers. If λ is not an even function, one has to consider

$$\begin{cases} \lambda_+(z,s) = \lambda(z,|s|) \\ \lambda_-(z,s) = \lambda(z,-|s|) \end{cases}$$

which are Young functions and the corresponding Orlicz spaces $L_{\lambda_\pm}$.

A.1. **Some definitions needed for stating Theorem 2.7**

The cone K_λ. It is the cone of all measurable functions u with a positive part u_+ in L_{λ_+} and a negative part u_- in L_{λ_-} : $K_\lambda = \{u$ measurable; $\exists a > 0, \int_{\mathcal{Z}} \lambda(au)\,dR < \infty\}$.

The cone $\widetilde{\mathcal{Y}}$. The $\sigma(K_\lambda, L_\pm)$-closure $\overline{A}$ of a set $A \subset K_\lambda$ is defined as follows: $u \in K_\lambda$ is in $\overline{A}$ if $u_\pm$ is in the $\sigma(L_{\lambda_\pm}, L_{\lambda_\pm^*})$-closure of $A_\pm = \{u_\pm; u \in A\}$. Clearly, $\overline{A}_\pm = \{u_\pm; u \in \overline{A}\}$. The cone $\widetilde{\mathcal{Y}} \subset \mathcal{X}^*$ is the extension of $\mathcal{Y}_o$ which is isomorphic to the $\sigma(K_\lambda, L_\pm)$-closure $\widetilde{T^*\mathcal{Y}_o}$ of $T^*\mathcal{Y}_o$ in K_λ in the sense that $T^*\widetilde{\mathcal{Y}} = \widetilde{T^*\mathcal{Y}_o}$.

The extended dual problem $(\widetilde{D}_C)$. The extended dual problem associated with (P_C) is

$$maximize \quad \inf_{x \in C \cap \mathcal{X}} \langle \omega, x \rangle - I_\gamma(\langle \omega, \theta \rangle), \quad \omega \in \widetilde{\mathcal{Y}}. \tag{$\widetilde{D}_C$}$$

Note that the dual bracket $\langle \omega, x \rangle$ is meaningful for each $\omega \in \widetilde{\mathcal{Y}}$ and $x \in \mathcal{X}$.

A.2. **Some definitions needed for stating Theorem 2.8**

Again, one needs to introduce several cones to state the extended dual problem $\overline{D}_C$.

Recall that there is a natural order on the algebraic dual space E^* of a Riesz vector space E which is defined by: $e^* \le f^* \Leftrightarrow \langle e^*, e \rangle \le \langle f^*, e \rangle$ for any $e \in E$ with $e \ge 0$. A linear form $e^* \in E^*$ is said to be *relatively bounded* if for any $f \in E$, $f \ge 0$, we have $\sup_{e:|e| \le f} |\langle e^*, e \rangle| < +\infty$. Although E^* may not be a Riesz space in general, the vector space E^b of all the relatively bounded linear forms on E is always a Riesz space. In particular, the elements of E^b admit a decomposition in positive and negative parts $e^* = e_+^* - e_-^*$.

The cone K''_λ. It is the cone of all relatively bounded linear forms $\zeta \in L'^b_{\lambda_\diamond}$ on $L'_{\lambda_\diamond}$ with a positive part ζ_+ whose restriction to $L'_{\lambda_+} \subset L'_{\lambda_\diamond}$ is in L''_{λ_+} and with a negative part ζ_- whose restriction to $L'_{\lambda_-} \subset L'_{\lambda_\diamond}$ is in L''_{λ_-} : $K''_\lambda = \{\zeta \in L'^b_{\lambda_\diamond}; \zeta_{\pm|L'_{\lambda_\pm}} \in L''_{\lambda_\pm}\}$. Note that $L'_{\lambda_\pm} \subset L'_{\lambda_\diamond}$.

A decomposition in K''_λ. Let ρ be any Young function. Translating decomposition (2.17) onto K''_λ leads to $K''_\lambda = [K_\lambda \oplus K^s_{\lambda^*}] \oplus K^{s'}_\lambda$ where one defines $K^s_{\lambda^*} = \{\zeta \in (L_{\lambda_\diamond^*} R)^b; \zeta_{\pm|L_{\lambda_\pm^*} R} \in L^s_{\lambda_\pm^*}\}$ and $K^{s'}_\lambda = \{\zeta \in L^{sb}_{\lambda_\diamond}; \zeta_{\pm|L^s_{\lambda_\pm}} \in L^{s'}_{\lambda_\pm}\}$.

Note that $L_{\lambda_\pm^*} R \subset L_{\lambda_\diamond^*} R$ and $L_{\lambda_\pm}^s \subset L_{\lambda_\diamond}^s$. With these cones in hand, the decomposition (2.17) holds for any $\zeta \in K_\lambda''$ with

$$\begin{cases} \zeta_1 = \zeta_1^a + \zeta_1^s \in K_\lambda \oplus K_{\lambda^*}^s = K_{\lambda^*}', \\ \zeta_2 \in K_\lambda^{s'}. \end{cases}$$

The set $\overline{\mathcal{Y}}$. The $\sigma(K_\lambda'', L_\pm')$-closure $\overline{A}$ of a set $A \subset K_\lambda''$ is defined as follows: $\zeta \in K_\lambda''$ is in $\overline{A}$ if $\zeta_\pm$ is in the $\sigma(L_{\lambda_\pm}'', L_{\lambda_\pm}')$-closure of $A_\pm = \{\zeta_\pm; \zeta \in A\}$. Clearly, $\overline{A}_\pm = \{\zeta_\pm; \zeta \in \overline{A}\}$. Let $\overline{T^*\mathcal{Y}_o}$ denote the $\sigma(K_\lambda'', L_\pm')$-closure of $T^*\mathcal{Y}_o$ in K_λ''.

Let D denote the $\sigma(K_\lambda^{s'}, L_\pm^s)$-closure of dom I_λ, that is $\zeta \in K_\lambda^{s'}$ is in D if and only if $\zeta_\pm$ is in the $\sigma(L_{\lambda_\pm}^{s'}, L_{\lambda_\pm}^s)$-closure of $\{u_\pm; u \in \operatorname{dom} I_\lambda\}$.

The set $\overline{\mathcal{Y}} \subset \mathcal{X}^*$ is the extension of $\mathcal{Y}_o$ which is isomorphic to $\overline{T^*\mathcal{Y}_o} \cap \{\zeta \in K_\lambda''; \zeta_1^s = 0, \zeta_2 \in D\}$ in the sense that

$$T^*\overline{\mathcal{Y}} = \overline{T^*\mathcal{Y}_o} \cap \{\zeta \in K_\lambda''; \zeta_1^s = 0, \zeta_2 \in D\}.$$

The extended dual problem $(\overline{\mathrm{D}}_C)$. The extended dual problem associated with $(\overline{\mathrm{P}}_C)$ is

$$maximize \quad \inf_{x \in C \cap \mathcal{X}} \langle \omega, x \rangle - I_\gamma([T^*\omega]_1^a), \quad \omega \in \overline{\mathcal{Y}}. \tag{$\overline{\mathrm{D}}_C$}$$

The exact statement of Theorem 2.8(c) is

Theorem A.1. *Under the assumptions of Theorem 2.8, we have*

(c) Let us denote $\hat{x} := T\hat{\ell}$. There exists $\bar{\omega} \in \overline{\mathcal{Y}}$ such that

$$\begin{cases} (a) & \hat{x} \in C \cap \operatorname{dom} \Gamma^* \\ (b) & \langle \bar{\omega}, \hat{x} \rangle_{\mathcal{X}^*, \mathcal{X}} \leq \langle \bar{\omega}, x \rangle_{\mathcal{X}^*, \mathcal{X}}, \forall x \in C \cap \operatorname{dom} \Gamma^* \\ (c) & \hat{\ell} \in \gamma_z'([T^*\bar{\omega}]_1^a) R + D^\perp([T^*\bar{\omega}]_2) \end{cases} \tag{A.1}$$

where

$$D^\perp(\eta) = \{k \in L_{\lambda_\diamond}^s; \forall h \in L_{\lambda_\diamond}, \eta + h \in D \Rightarrow \langle h, k \rangle \leq 0\}$$

is the outer normal cone of D at η.

$T^*\bar{\omega}$ is in the $\sigma(K_\lambda'', L_\pm')$-closure of $T^*(\{y \in \mathcal{Y}_o; \int_{\mathcal{Z}} \lambda(\langle y, \theta \rangle) \, \mathrm{d}R < \infty\})$ and there exists some $\tilde{\omega} \in \mathcal{X}_o^*$ such that

$$[T^*\bar{\omega}]_1^a = \langle \tilde{\omega}, \theta(\cdot) \rangle_{\mathcal{X}_o^*, \mathcal{X}_o}$$

is a measurable function in the strong closure of $T^*(\{y \in \mathcal{Y}_o; \int_{\mathcal{Z}} \lambda(\langle y, \theta \rangle) \, \mathrm{d}R < \infty\})$ in K_λ: the set of all $u \in K_\lambda$ such that $u_\pm$ is in the strong closure of $T^*(\{y \in \mathcal{Y}_o; \int_{\mathcal{Z}} \lambda(\langle y, \theta \rangle) \, \mathrm{d}R < \infty\})_\pm$ in $L_{\lambda_\pm}$.

Furthermore, $\hat{\ell} \in L_{\lambda_\diamond}'$ and $\bar{\omega} \in \overline{\mathcal{Y}}$ satisfy (A.1) if and only if $\hat{\ell}$ solves $(\overline{\mathrm{P}}_C)$ and $\bar{\omega}$ solves $(\overline{\mathrm{D}}_C)$.

The dual bracket $\langle [T^*\bar{\omega}]_2, \hat{\ell}^s \rangle$ in Theorem 2.8(d.3) is intended to be $\langle [T^*\bar{\omega}]_2, \hat{\ell}^s \rangle_{K_\lambda^{s'}, K_\lambda^s}$.

A.3. **A complement to Theorem 4.1**

As an easy corollary of Theorem A.1 and the proof of Theorem 4.1 we obtain the following

Proposition A.2 (complement to Thm. 4.1). *For each $\omega_\diamond$ satisfying (4.1) at Theorem 4.1, $\langle \omega_\diamond, \theta(\cdot) \rangle$ is in the strong closure of $T^*\operatorname{dom} \Gamma$ in K_λ: the set of all $u \in K_\lambda$ such that $u_\pm$ is in the $\| \cdot \|_{\lambda_\pm}$-closure of $\{\langle y, \theta \rangle_\pm; y \in \operatorname{dom} \Gamma\}$ in $L_{\lambda_\pm}$, and there exists $\bar{\omega} \in \overline{\mathcal{Y}}$ solution of $(\overline{\mathrm{D}}_C)$ such that $[T^*\bar{\omega}]_1^a = \langle \omega_\diamond, \theta(\cdot) \rangle$.*

Recall that Γ is defined at (2.12).

Appendix B. Proof of Theorem 7.1

This proof is an easy variation on [9] (Sect. 7.3). It is given for the reader's convenience. Theorem 7.1 is a restatement of Propositions B.1, B.3, B.4 and B.6 below.

We begin with the proof for X_n. As T is continuous and I is a good rate function, J is also a good rate function. Let us first state the upper bound of a conditional LDP.

Proposition B.1. *Under the assumptions of Theorem 7.1, for all closed subset F of $\mathcal{X}_o$, we have*

$$\limsup_{\delta \to 0} \limsup_{n \to \infty} \frac{1}{n} \log \mathbb{P}(X_n \in F \mid X_n \in C_\delta) \leq -J_C(F)$$

where

$$J_C(x) := \begin{cases} J(x) - J(C) & \text{if } x \in C \\ +\infty & \text{if } x \notin C. \end{cases}$$

Proof. Clearly, for all measurable set B and all $n \geq 1$, $\delta > 0$, we have $\frac{1}{n} \log \mathbb{P}(X_n \in B \mid X_n \in C_\delta) = \frac{1}{n} \log \mathbb{P}(X_n \in B \cap C_\delta) - \frac{1}{n} \log \mathbb{P}(X_n \in C_\delta)$. Hence,

$$
\begin{aligned}
\limsup_n \frac{1}{n} \log \mathbb{P}(X_n \in F \mid X_n \in C_\delta) &\leq \limsup_n \frac{1}{n} \log \mathbb{P}(X_n \in F \cap \mathrm{cl}\, C_\delta) - \liminf_n \frac{1}{n} \log \mathbb{P}(X_n \in \mathrm{int}\, C_\delta) \\
&\leq -J(F \cap C_\delta) + J(\mathrm{int}\, C_\delta) \\
&\leq -J(F \cap C_\delta) + J(C)
\end{aligned}
$$

where the last inequality follows from the assumption (B_C^3). We complete the proof with the following lemma. $\square$

Lemma B.2. *For any closed set F, $\lim_\delta J(F \cap C_\delta) = J(F \cap C) \in [0, \infty]$.*

Proof. As $C \subset C_\delta$, for all $\delta > 0$, we have $J(F \cap C_\delta) \leq J(F \cap C)$. Since C_δ is decreasing, $J(F \cap C_\delta)$ is nondecreasing and $\lim_\delta J(F \cap C_\delta) = \sup_\delta J(F \cap C_\delta) \in [0, \infty]$. If $\sup_\delta J(F \cap C_\delta) = \infty$, the inequality $J(F \cap C_\delta) \leq J(F \cap C)$ leads to the desired result.

Now, let us suppose that $\sup_\delta J(F \cap C_\delta) < \infty$. As $F \cap C_\delta$ is closed and J is inf-compact, for any δ there exists $x_\delta \in F \cap C_\delta$ such that $J(x_\delta) = J(F \cap C_\delta)$ and we can extract a converging subsequence $x_k \to x_*$. Because the C_δ's are decreasing, we get $\cap_\delta C_\delta = \cap_k C_{\delta_k}$ and $\lim_\delta J(F \cap C_\delta) = \lim_k J(x_k)$. More, $x_* \in F \cap (\cap_k C_{\delta_k}) = F \cap C$ and as J is lsc: $\lim_k J(x_k) \geq J(x_*) \geq J(F \cap C)$. Therefore, $\lim_\delta J(F \cap C_\delta) \geq J(F \cap C)$ which completes the proof. $\square$

Let us state the lower bound corresponding to Proposition B.1.

Proposition B.3. *If the assumption (B_C^3) on the conditioning event is restricted to (3b):*

$$C \subset \mathrm{int}\, C_\delta, \forall \delta > 0 \tag{B.1}$$

then, for all open subset G of $\mathcal{X}_o$, we have

$$\liminf_{\delta \to 0} \liminf_{n \to \infty} \frac{1}{n} \log \mathbb{P}(X_n \in G \mid X_n \in C_\delta) \geq -J_C(G).$$

Proof. For all $\delta > 0$,

$$
\begin{aligned}
\liminf_n \frac{1}{n} \log \mathbb{P}(X_n \in G \mid X_n \in C_\delta) &\geq \liminf_n \frac{1}{n} \log \mathbb{P}(X_n \in G \cap \mathrm{int}\, C_\delta) - \limsup_n \frac{1}{n} \log \mathbb{P}(X_n \in C_\delta) \\
&\geq -J(G \cap \mathrm{int}\, C_\delta) + J(C_\delta) \\
&\geq -J(G \cap C) + J(C_\delta).
\end{aligned}
$$

We conclude with Lemma B.2.   □

Let us recall that $\mathcal{H} = \operatorname{argmin} J_C$ is the set of the minimizers of J on C. As C is closed and J is inf-compact, $\mathcal{H}$ is a compact set. As an immediate corollary of Proposition B.1, we have the following conditional LLN which is the part of the statement of Theorem 7.1 concerning X_n.

Proposition B.4. *For all open subset H of $\mathcal{X}_o$ such that $\mathcal{H} \subset H$, we have*

$$\limsup_{\delta \to 0} \limsup_{n \to \infty} \frac{1}{n} \log \mathbb{P}(X_n \notin H \mid X_n \in C_\delta) < 0.$$

Let us now have a look at L_n. We are interested in the asymptotic behavior of $\mathbb{P}(L_n \in \cdot \mid TL_n \in C_\delta)$ with $C_\delta \subset \mathcal{X}_o$. Let us denote $A_\delta := T^{-1}(C_\delta) = \{\ell \in \mathcal{L}_o; T\ell \in C_\delta\}$ and $A = T^{-1}C$. It is useful to state the assumptions on the C_δ's rather than on the A_δ's. In fact, we have the following transfer result.

Lemma B.5. *We assume that T is continuous.*
  (a) *If C is closed and $J(C) = J(\operatorname{int} C)$, then A is closed and $I(A) = I(\operatorname{int} A)$.*
  (b) *If $C := \cap_\delta C_\delta \subset \operatorname{int} C_\delta$ for all $\delta > 0$, then $A = \cap_\delta \operatorname{cl} A_\delta$ and $A \subset \operatorname{int} A_\delta$ for all $\delta > 0$.*

Proof. Since T is continuous, for any $A' = T^{-1}C'$, we have: $T^{-1}(\operatorname{int} C') \subset \operatorname{int} A' \subset A' \subset \operatorname{cl} A' \subset T^{-1}(\operatorname{cl} C')$.

Let us begin with (a). As C is closed, so is A. For any $A = T^{-1}C$, we have $I(A) = \inf\{\Phi^*(\ell); T\ell \in C\} = \inf_{x \in C} \inf\{\Phi^*(\ell); T\ell = x\} = \inf_{x \in C} J(x) = J(C)$. Hence, $I(A) = J(C) = J(\operatorname{int} C)$ (by hypothesis) $= I(T^{-1}(\operatorname{int} C) \geq I(\operatorname{int} A)$, since $T^{-1}(\operatorname{int} C) \subset \operatorname{int} A$. But the converse inequality: $I(A) \leq I(\operatorname{int} A)$ is clear.

Let us prove (b). We have: $\cap_\delta \operatorname{cl} A_\delta \subset T^{-1}(\cap_\delta C_\delta) := A := T^{-1}(C) \subset T^{-1}(\cap_\delta \operatorname{int} C_\delta)$ (by hypothesis) $\subset \cap_\delta \operatorname{int} A_\delta \subset \cap_\delta \operatorname{cl} A_\delta$. This proves that all these sets are equal, and in particular that $A = \cap_\delta \operatorname{cl} A_\delta$. On the other hand, as for any $\delta > 0$, $C \subset \operatorname{int} C_\delta$, we have $A = T^{-1}(C) \subset T^{-1}(\operatorname{int} C_\delta) \subset \operatorname{int} A_\delta$.   □

Let us recall that $\mathcal{G}$ is the set of the minimizers of I on A. By the above Lemma B.5, in the situation of the L_n's, Proposition B.4 becomes the following

Proposition B.6. *Under our general assumptions, for all open subset G of $\mathcal{L}_o$ such that $\mathcal{G} \subset G$, we have*

$$\limsup_{\delta \to 0} \limsup_{n \to \infty} \frac{1}{n} \log \mathbb{P}(L_n \notin G \mid TL_n \in C_\delta) < 0.$$

Note that by Lemma B.5, the A_δ's share the same properties as the C_δ's. In particular, $I(\operatorname{int} A_\delta) \leq I(A)$ also holds for them.

REFERENCES

[1] R. Azencott, Grandes déviations et applications, in *École d'Eté de Probabilités de Saint-Flour VIII* (1978).
[2] J.M. Borwein and A.S. Lewis, Strong rotundity and optimization. *SIAM J. Optim.* **1** (1994) 146–158.
[3] C. Boucher, R.S. Ellis and B. Turkington, Spatializing random measures: doubly indexed processes and the large deviation principle. *Ann. Probab.* **27** (1999) 297–324.
[4] I. Csiszár, *I*-divergence geometry of probability distributions and minimization problems. *Ann. Probab.* **3** (1975) 146–158.
[5] I. Csiszár, Sanov property, generalized *I*-projection and a conditional limit theorem. *Ann. Probab.* **12** (1984) 768–793.
[6] I. Csiszár, Generalized projections for non-negative functions. *Acta Math. Hungar.* **68** (1995) 161–185.
[7] I. Csiszár, F. Gamboa and E. Gassiat, MEM pixel correlated solutions for generalized moment and interpolation problems. *IEEE Trans. Inform. Theory* **45** (1999) 2253–2270.
[8] D. Dacunha-Castelle and F. Gamboa, Maximum d'entropie et problème des moments. *Ann. Inst. H. Poincaré. Probab. Statist.* **26** (1990) 567–596.
[9] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*. Second edition. Appl. Math. **38**. Springer-Verlag (1998).
[10] U. Einmahl and J. Kuelbs, Dominating points and large deviations for random vectors. *Probab. Theory Relat. Fields* **105** (1996) 529–543.

[11] R.S. Ellis, J. Gough and J.V. Puli, The large deviations principle for measures with random weights. *Rev. Math. Phys.* **5** (1993) 659–692.

[12] F. Gamboa and E. Gassiat, Bayesian methods and maximum entropy for ill-posed inverse problems. *Ann. Statist.* **25** (1997) 328–350.

[13] H. Gzyl, *The Method of Maximum Entropy*. World Scientific (1994).

[14] J. Kuelbs, Large deviation probabilities and dominating points for open convex sets: nonlogarithmic behavior. *Ann. Probab.* **28** (2000) 1259–1279.

[15] C. Léonard, Large deviations for Poisson random measures and processes with independent increments. *Stoch. Proc. Appl.* **85** (2000) 93–121.

[16] C. Léonard, Convex minimization problems with weak constraint qualifications. *J. Convex Anal.* **17** (2010) 321–348.

[17] C. Léonard, Minimization of energy functionals applied to some inverse problems. *J. Appl. Math. Optim.* **44** (2001) 273–297.

[18] C. Léonard, Minimizers of energy functionals under not very integrable constraints. *J. Convex Anal.* **10** (2003) 63–88.

[19] C. Léonard, Minimization of entropy functionals. *J. Math. Anal. Appl.* **346** (2008) 183–204.

[20] C. Léonard and J. Najim, An extension of Sanov's theorem: application to the Gibbs conditioning principle. *Bernoulli* **8** (2002) 721–743.

[21] J. Najim, A Cramér type theorem for weighted random variables. *Electron. J. Probab.* **7** (2002) 1–32.

[22] P. Ney, Dominating points and the asymptotics of large deviations for random walks on R^d. *Ann. Probab.* **11** (1983) 158–167.

[23] P. Ney, Convexity and large deviations. *Ann. Probab.* **12** (1984) 903–906.

[24] M.M. Rao and Z.D. Ren, *Theory of Orlicz spaces*, Pure Appl. Math. **146**. Marcel Dekker, Inc. (1991).

[25] R.T. Rockafellar, Integrals which are convex functionals. *Pacific J. Math.* **24** (1968) 525–539.

[26] R.T. Rockafellar, *Conjugate Duality and Optimization*, volume 16 of *Regional Conf. Series in Applied Mathematics*. SIAM, Philadelphia (1974).

[27] R.T. Rockafellar and R. Wets, *Variational Analysis*, in *Grundlehren der Mathematischen Wissenschaften*, volume 317. Springer (1998).

[28] A. Schied, Cramér's condition and Sanov's theorem. *Statist. Probab. Lett.* **39** (1998) 55–60.

[29] C.R. Smith, G.J. Erickson and P.O. Neudorfer (Eds.). *Maximum Entropy and Bayesian Methods, Proc. of 11th Int. Workshop on Maximum Entropy and Bayesian Methods of Statistical Analysis, Seattle, 1991*. Kluwer.

[30] D.W. Stroock and O. Zeitouni, Microcanonical distributions, Gibbs states and the equivalence of ensembles, in *Festchrift in Honour of F. Spitzer*, edited by R. Durrett and H. Kesten. Birkhaüser (1991) 399–424.

Index