ECOLE POLYTECHNIQUE

CENTRE DE MATHÉMATIQUES APPLIQUÉES

UMR CNRS 7641

91128 PALAISEAU CEDEX (FRANCE). Tél: 01 69 33 41 50. Fax: 01 69 33 30 11

http://www.cmap.polytechnique.fr/

Dominating Points and Entropic Projections

Christian Léonard

R.I. N^0 525 *March 2004*

DOMINATING POINTS AND ENTROPIC PROJECTIONS

CHRISTIAN LÉONARD

Abstract. Some Conditional Laws of Large Numbers (CLLN) are related to minimization problems: the limit of the CLLN is the minimizer of a large deviation rate function on the limiting conditioning set. When the CLLN is concerned with empirical means, the minimizer is called a dominating point; if it is concerned with empirical measures, it is called an entropic projection.
CLLNs are obtained both for empirical means and measures with independent random weights. By means of convex conjugate duality, one obtains dual equalities and dual representations of the minimizers: the dominating points and the entropic projections. For some convex conditioning events, it may happen that usual integral representations of the dominating point fail: no entropic projection exists. This phenomenon is clarified by introducing extended minimization problems the minimizers of which may not be measures anymore. It appears that in some situations, the generalized entropic projections discovered by Csiszár are the measure component of these extended minimizers. The important case of relative entropy is studied in details.

1. Introduction

Let $\{X_n\}_{n\geq 1}$ be a sequence of random elements which obeys a Large Deviation Principle (LDP) in a topological vector space $\mathcal{X}$ with rate function J. Let us consider the sequence of conditioning events $\{X_n \in C\}$ where C is a subset of $\mathcal{X}$. Under some assumptions, one can prove the following Conditional Law of Large Numbers (CLLN): $\mathbb{P}(X_n \in \cdot \mid X_n \in C)$ converges weakly as n tends to infinity to the Dirac unit mass $\delta_{\bar{x}}$ at $\bar{x} \in C$ where $\bar{x}$ is the solution to the minimization problem

$$\textit{minimize } J(x) \textit{ subject to } x \in C. \tag{1.1}$$

In the situation where J is the rate function of Cramér's theorem, Ney [25], [26] called $\bar{x}$ a predominating point of C for the rate function J. A *dominating point* shares additional properties, in particular it admits an integral representation in terms of a dual parameter, see Definition 6.1 for more details.

Let $\{L_n\}$ be a family of random measures on a measurable space Ω which obeys a LDP with rate function I. Let us consider a sequence of conditoning events $\{L_n \in \mathcal{C}\}$ where $\mathcal{C}$ is a subset of $\mathcal{M}(\Omega)$: the space of all signed measures on Ω. In this situation, the predominating point $\bar{\ell}$ of $\mathcal{C}$ for the rate function I is the solution of the minimization problem

$$\textit{minimize } I(\ell) \textit{ subject to } \ell \in \mathcal{C}$$

Date: March 2004.

1991 *Mathematics Subject Classification.* Primary: 60F10, 60F99, 60G57 Secondary: 46N10.

Key words and phrases. Conditional laws of large numbers, random measures, large deviations, entropy, convex optimization.

and one can expect that $\mathbb{P}(L_n \in \cdot \mid L_n \in \mathcal{C})$ converges weakly as n tends to infinity to $\delta_{\bar{\ell}}$. When L_n is the empirical measure of an iid sequence, then I is the relative entropy $I(\cdot \mid R)$ with respect to the common law R of the sampled random variables. This minimization problem is studied extensively by Csiszár in [6] and [7] where the minimizer $\bar{\ell}$ is called the *I-projection* of R on $\mathcal{C}$. Csiszár also proves that when $\mathcal{C}$ is a convex set, any minimizing sequence converges in variation to some probability measure ℓ_* which may not belong to $\mathcal{C}$, with $I(\ell_* \mid R) \leq \inf_{\ell \in \mathcal{C}} I(\ell \mid R)$ and possibly a strict inequality. Csiszár calls this ℓ_* the *generalized I-projection* of R on $\mathcal{C}$. Later in [8], he extends this notion to a larger class of convex functionals I on $\mathcal{M}(\Omega)$, which will be called *entropic projections* and *generalized entropic projections* in the present paper, see Definition 4.2 below. The form of these convex functionals is

$$I(\ell) = \int_\Omega \lambda^* \left(\frac{d\ell}{dR}\right) dR \tag{1.2}$$

if ℓ is absolutely continuous with respect to R and $+\infty$ otherwise, where $\lambda^* : \mathbb{R} \to [0, \infty]$ is a convex function. Let us consider the *empirical measures with random weights*

$$L_n = \frac{1}{n} \sum_{i=1}^n W_i^n \delta_{\omega_i^n} \in \mathcal{M}(\Omega), n \geq 1$$

where the deterministic sequence of probability measures $\frac{1}{n} \sum_{i=1}^n \delta_{\omega_i^n} \in \mathcal{P}(\Omega)$ is assumed to converge as n tends to infinity to the reference probability measure R and the random weights W_i^n are assumed to be iid. It is well known that if λ^* is the Cramér transform of the common law of the W_i^n's, I is the rate function of the LDP for $\{L_n\}$, see for instance the articles of Ellis, Gough and Puli [11] and Boucher, Ellis and Turkington [3] for motivations in statistical physics and the article of Gamboa and Gassiat [14] for motivations in statistics.

Suppose now that X_n and L_n are linked together by the relation $X_n = \int_\Omega \theta(\omega)\, L_n(d\omega)$ where $\theta : \Omega \to \mathcal{X}$ is a vector-valued measurable mapping. Under some regularity conditions, the contraction principle allows us to derive the LDP for $\{X_n\}$ from the LDP for $\{L_n\}$ and the corresponding rate functions satisfy: $J(x) = \inf\{I(\ell); \ell \in \mathcal{M}(\Omega), \langle \theta, \ell \rangle = x\}$. This suggests to consider the particular constraint set $\mathcal{C} = \{\ell \in \mathcal{M}(\Omega); \langle \theta, \ell \rangle \in C\}$ and the minimization problem

$$\textit{minimize } \int_\Omega \lambda^* \left(\frac{d\ell}{dR}\right) dR \textit{ subject to } \langle \theta, \ell \rangle \in C, \quad \ell \in \mathcal{M}(\Omega). \tag{1.3}$$

From the previous arguments, one can expect that the empirical measures with random weights satisfy the following CLLN

$$\mathbb{P}\left(\frac{1}{n} \sum_{i=1}^n W_i^n \delta_{\omega_i^n} \in \cdot \quad \Big| \quad \frac{1}{n} \sum_{i=1}^n W_i^n \theta(\omega_i^n) \in C\right) \textit{ converges weakly to } \delta_{\bar{\ell}} \tag{1.4}$$

where $\bar{\ell}$ is the solution of the minimization problem (1.3).

Let us assume that C is a convex set, so that (1.1) and (1.3) are convex minimization problems. Introducing a vector space $\mathcal{Y}$ in dual pairing with $\mathcal{X}$ allows us to state the following unconstrained dual problem associated with the primal problem (1.3)

$$\textit{maximize } \inf_{x \in C} \langle y, x \rangle - \int_\Omega \lambda(\langle y, \theta \rangle)\, dR, \quad y \in \mathcal{Y} \tag{1.5}$$

where λ and λ^* are convex conjugates, that is λ is the log-Laplace transform of W_i^n. At least formally, the classical approach to convex optimization as developped in the monograph [30] by Rockafellar suggests that under some regularity assumptions, the following *dual equalities*

$$\begin{aligned}
\inf\{I(\ell); \langle\theta,\ell\rangle = x, \ell\in\mathcal{M}(\Omega)\} &= \sup_{y\in\mathcal{Y}}\left\{\langle y,x\rangle - \int_\Omega \lambda(\langle y,\theta\rangle)\,dR\right\}, \quad x\in\mathcal{X}\\
\inf\{I(\ell); \langle\theta,\ell\rangle \in C, \ell\in\mathcal{M}(\Omega)\} &= \sup_{y\in\mathcal{Y}}\left\{\inf_{x\in C}\langle y,x\rangle - \int_\Omega \lambda(\langle y,\theta\rangle)\,dR\right\}
\end{aligned}$$

should hold (note that the contraction principle leads us to $\inf\{I(\ell); \langle\theta,\ell\rangle = x, \ell\in\mathcal{M}(\Omega)\} = J(x)$ and $\inf\{I(\ell); \langle\theta,\ell\rangle \in C, \ell\in\mathcal{M}(\Omega)\} = \inf_{x\in C} J(x)$) and the minimizers $\bar\ell$ and $\bar x$ of (1.3) and (1.1) should be represented as

$$\begin{aligned}
\bar\ell(d\omega) &= \lambda'(\langle\bar y,\theta(\omega)\rangle)\,R(d\omega)\\
\bar x &= \langle\theta,\bar\ell\rangle = \int_\Omega \theta(\omega)\lambda'(\langle\bar y,\theta(\omega)\rangle)\,R(d\omega)
\end{aligned}$$

where $\bar y$ is some maximizer of (1.5). These identities are called *dual representations* of the minimizers.

1.1. **Presentation of the results.** The aim of this paper is to prove CLLNs of the type of (1.4) and to obtain dual equalities together with dual representations of the dominating points $\bar x$ and the entropic projections $\bar\ell$ as above.

A CLLN for empirical measures with random weights in the spirit of (1.4) is stated at Theorem 2.19 in Section 2. To be more precise, our results hold for independent random weights W_i^n which may not be identically distributed. The log-Laplace $\lambda(\omega,\cdot)$ of W_i^n depends on ω_i^n and the rate function of the LDP for $\{L_n\}$ is $I(\ell) = \int_\Omega \lambda^*\left(\omega, \frac{d\ell}{dR}(\omega)\right) R(d\omega)$. Moreover, our assumptions allow us to consider conditioning sets $C\subset\mathcal{X}$ such that $\mathbb{P}(\langle\theta,L_n\rangle\in C)$ may vanish. The CLLNs are in terms of $\lim_{\delta\to0}\lim_{n\to\infty}\mathbb{P}(L_n\in\cdot\mid\langle\theta,L_n\rangle\in C_\delta)$ where C_δ is a blowup of C such that $\mathbb{P}(\langle\theta,L_n\rangle\in C_\delta)>0$ for all n and δ.

The proofs of the dual equalities and representations of the minimizers are based on the classical conjugate duality approach applied in the convenient functional analysis framework obtained by means of pertinent Orlicz spaces. These general results are obtained in another article of the author [22].

It appears that if the constraint mapping θ is integrable enough (see the "*good constraints*" assumption (3.3)), the minimization problem (1.3) is attained in $\mathcal{M}(\Omega)$ provided that $\inf\{I(\ell);\ell,\int_\Omega\theta\,d\ell\in C\}<\infty$. The corresponding results about dual equalities and representations of $\bar x$ and $\bar\ell$ are collected at Theorem 3.7 and Proposition 4.3.

But it may happen that when θ isn't integrable enough (see the "*bad constraints*" assumption (3.6)), although $\inf\{I(\ell);\ell,\int_\Omega\theta\,d\ell\in C\}<\infty$, the minimization problem (1.3) is not attained in $\mathcal{M}(\Omega)$ anymore. Nevertheless, as proved by Csiszár in [8], a generalized entropic projection still exists in $\mathcal{M}(\Omega)$. Considering an extended primal minimization problem associated with the same maximization dual problem (see $(\bar P)$ at Section 4), one is able to prove at Section 4 that the minimizer of the extended problem is the sum of a measure $\bar\ell^a$ which is absolutely continuous with respect to R and a singular element $\bar\ell^s$ which is not even a measure in the general case. It is also proved that if in addition the constraint is finite dimensional, the absolutely continuous component $\bar\ell^a$ of this minimizer is the generalized entropic projection $\ell_* = \bar\ell^a$ and that the singular component $\bar\ell^s$ is responsible for the possible gap: $I(\ell_*) < \inf\{I(\ell); \ell \in \mathcal{M}(\Omega), \langle \theta, \ell\rangle \in C\} = I(\ell_*) + \bar I(\bar\ell^s)$, where $\bar I$ is some extension of I. These results are collected at Theorems 4.5 and 4.7.

The case of the relative entropy is of special interest. It is associated with CLLNs for the empirical measure of an R-iid sample. In this paper, we shall take advantage of the fact that it also corresponds to CLLNs for empirical measures with Poisson(1)-distributed random weights with an additional unit mass constraint: $\langle \mathbf{1}, \ell\rangle = 1$. In this situation, $\lambda(s) = e^s - 1$: the log-Laplace transform of a Poisson(1) law. It is studied in details at Section 5. Our representation results are stated at Propositions 5.3 and 5.5. The main result of Section 5 states at Theorem 5.9 that with $\mathcal{X}$ a Banach space, for a quite general constraint mapping θ and under the "bad constraint" assumption that the generalized I-projection is the common absolutely continuous part of the minimizers of the extended relative entropy (see (5.1)). Together with Theorem 4.7, these results shed light on the surprising phenomenon of *generalized* entropic projections.

At Section 6 it is shown that some predominating points may fail to be dominating when the convex conjugate $\Lambda^*(x)$ of $\Lambda(y) = \int_\Omega \lambda(\omega, \langle y, \theta(\omega)\rangle)\, R(d\omega)$ admits a nondegenerate recession function. This result is stated at Theorem 6.6.

1.2. **About the literature.** Conditional LLNs are well known. They already appear in the fundamental works about Large Deviations of Freidlin and Wentzell [13] and Azencott [2]. CLLNs for empirical measures of iid samples are obtained by Csiszár in [7]. In the context of statistical physics, they are sometimes called Gibbs Conditoning Principle. One can see for instance the paper [31] by Stroock and Zeitouni and more recently (Léonard and Najim, [23]). In the present paper, the LDP for empirical measures with random weights stated at Theorem 2.16 is obtained under "good" (i.e. strong) integrability assumptions. Such a LDP with weaker integrability assumptions but stronger regularity requirements is obtained by Najim in [24].

Functionals of the type of (1.2) are sometimes called λ^*-entropies, see for instance the paper [32] by Teboulle and Vajda. They are studied in the spirit of the present paper by Csiszár in [8] with other methods of proof. Theorem 3.7 and Proposition 4.3 extend similar results of Csiszár ([8], Thm 3 and its corollary).

The dual equalities in Theorems 3.7 and 4.5 already appear in the author's paper [20].

Generalized entropic projections have been discovered and studied by Csiszár in [7] and [8]. Our representation of the generalized entropic projections under finitely many constraints stated at Theorem 4.7 extends Csiszár's result ([8], Theorem 4). The interpretation of the generalized entropic projection as the absolutely continuous component of the minimizers $\bar\ell$ of an extended minimization problem is a new result.

The special case of the relative entropy has been extensively studied by Csiszár ([6], [7]). The representation of the dominating points in Proposition 5.3 has already been obtained for C with a nonempty topological interior in $\mathbb{R}^d$ by Ney in [26] and in a Banach space setting by Einmahl and Kuelbs in [10]. Proposition 5.3 extends these results. Proposition 5.5 also extends corresponding results of (Kuelbs, [17]) which are obtained in a Banach space setting with a "nonempty interior" assumption. The representation of the I-projection is obtained by means of a very different proof by Csiszár in [6] and ([7], Thm 3).

Integral representations of dominating points rely essentially on the existence of a supporting hyperplane shared by C and some level set of $J = \Lambda^*$. This existence is a consequence of Hahn-Banach theorem. In (Einmahl and Kuelbs, [10]) and (Kuelbs, [17]), the hypotheses of Hahn-Banach theorem are fulfilled by assuming that C has a nonempty topological interior in a Banach space. In order to get rid of this restrictive assumption in the present paper, made-to-measure topologies are used to insure that the level sets of $J = \Lambda^*$ have nonempty interiors in some sense. This strategy is developped in [22] by means of Orlicz spaces naturally associated with λ and λ^*. The main results of [22] that will be used in the present article are collected in Appendix B for the convenience of the reader.

In (Kuelbs, [17]) and [21], integral representations of dominating points are obtained under bad integrability assumptions. Theorem 6.6 complements and extends these results.

Contents

2. Conditional laws of large numbers for empirical measures with random weights

A CLLN for empirical measures with random weights is stated at Theorem 2.19. Its proof relies upon two results: a general result (Theorem 2.3) which allows us to derive a CLLN from a LDP, and a LDP for empirical measures with random weights which is stated at Theorem 2.16. Examples of interesting conditioning events are also described at Section 2.2. They correspond to moment and marginal constraints.

2.1. **Conditional laws of large numbers.** Let $\{L_n\}$ be a sequence of random vectors in some vector space $\mathcal{L}$. Let $T : \mathcal{L} \to \mathcal{X}$ be a linear operator with values in another vector space $\mathcal{X}$. 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 $\mathcal{X}$. It appears that this type of CLLN is connected with large deviations. We assume that $\{L_n\}$ obeys the LDP in $\mathcal{L}$ endowed with some topology and the associated Borel σ-field, with a good rate function I. 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 is a blowup that tends to C as δ tends to zero, with $\mathbb{P}(TL_n \in C_\delta) > 0$ for all n and δ.

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

$$X_n \stackrel{\Delta}{=} TL_n$$

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

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

Before stating a result about CLLNs at Theorem 2.3, we have to fix some assumptions on $\{L_n\}$ and the conditioning events.

Assumptions on $\{L_n\}$. *The sequence $\{L_n\}$ obeys the LDP in $\mathcal{L}$ with a* good *rate function I. This means that I is inf-compact.*

As a convention, one writes $J(C)$ for $\inf_{x\in C} J(x)$.

Assumptions on the conditioning events. This framework is based on (Stroock and Zeitouni, [31]) and (Dembo and Zeitouni, [9], Section 7.3).

(a) *The linear operator $T : \mathcal{L} \to \mathcal{X}$ is continuous.*
(b) $J(C) < \infty$.
(c) *The set C is closed, it is the limit as δ decreases to 0: $C \stackrel{\Delta}{=} \cap_\delta \mathrm{cl}\, C_\delta$, of the closures of a nonincreasing family of Borel sets C_δ in $\mathcal{X}$ such that for all $\delta > 0$ and all $n \geq 1$, $\mathbb{P}(X_n \in C_\delta) > 0$*
(d) *and one of the following statements*
   (1) *$C_\delta = C$ for all $\delta > 0$ and $J(\mathrm{int}\, C) = J(C)$, or*
   (2) *$C \subset \mathrm{int}\, C_\delta$ for all $\delta > 0$.*
   is fulfilled.

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

$$\text{minimize } I(\ell) \text{ subject to } T\ell \in C, \quad \ell \in \mathcal{L}. \tag{2.1}$$

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

$$\text{minimize } J(x) \text{ subject to } x \in C, \quad x \in \mathcal{X}. \tag{2.2}$$

We can now state a result about CLLNs.

Theorem 2.3 (CLLN). *For all open subset G of $\mathcal{L}$ such that $\mathcal{G} \subset G$ and all open subset H of $\mathcal{X}$ 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 is strictly convex, we have the CLLNs:

$$\begin{aligned}\lim_\delta\lim_n \mathbb{P}(L_n \in \cdot \mid TL_n \in C_\delta) &= \delta_{\bar\ell}\\ \lim_\delta\lim_n \mathbb{P}(X_n \in \cdot \mid X_n \in C_\delta) &= \delta_{\bar x}\end{aligned}$$

where the limits are understood with respect to the usual weak topologies of probability measures and $\bar\ell$ is the unique solution to the convex minimization problem (2.1) and $\bar x = T\bar\ell$ is the unique solution to (2.2).

Let us note that if I is inf-compact and strictly convex, then J is also strictly convex. The proof of this theorem is postponed to Section 7.

2.2. **Random measures.** Let Ω be a space endowed with a σ-field $\mathcal{A}$. We denote $\mathcal{P}(\Omega)$ the set of all probability measures and $\mathcal{M}(\Omega)$ the space of all *signed* measures on $(\Omega, \mathcal{A})$. In this paper, a measure will always be a signed measure.
Let $\mathcal{V}$ be a space of measurable functions $v : \Omega \to \mathbb{R}$. We shall consider the subspace of its algebraic dual space: $\mathcal{V}^\sharp$, which consists of linear forms on $\mathcal{V}$ which act as measures:

$$\mathcal{M}^{\mathcal{V}} \triangleq \{\ell \in \mathcal{V}^\sharp; \exists Q_\ell \in \mathcal{M}(\Omega), \langle \ell - Q_\ell, v\rangle = 0, \forall v \in \mathcal{V} \text{ and } \int_\Omega v\, d|Q_\ell| < \infty, \forall v \in \mathcal{V}, v \geq 0\} \quad (2.4)$$

As a typical example, let us consider $L_n = \frac{1}{n}\sum_{i=1}^n \delta_{Z_i} \in \mathcal{P}(\Omega)$: the empirical measure of an iid sequence (Z_i) of Ω-valued random variables. Suppose in addition that Ω is a vector space, take $\mathcal{X} := \Omega$ and consider the operator $T : \ell \in \mathcal{M}^{\mathcal{V}} \mapsto$ "$\int_\Omega \omega\, \ell(d\omega)$" $\in \Omega$ where $\mathcal{M}^{\mathcal{V}} \subset \mathcal{M}(\Omega)$ is the space of all measures ℓ such that the integral $\int_\Omega \omega\, \ell(d\omega)$ makes sense. This corresponds to the space $\mathcal{V}$ of all "sublinear" measurable functions on Ω. Then, $TL_n = X_n = \frac{1}{n}\sum_{i=1}^n Z_i$ is the empirical mean and the large deviations for $\{L_n\}$ and $\{X_n\}$ are described by *Sanov*'s and *Cramér*'s theorems.

Constraint operators and moment functions. Let us go back to a general setting for random measures. It is convenient to describe the constraint operator T by means of some adjoint operator T^T. To do it, one needs to introduce dual spaces. Clearly, $\mathcal{V}$ and $\mathcal{M}^{\mathcal{V}}$ are in separating duality.
We also assume that $\mathcal{X}$ is in separating duality with another vector space $\mathcal{Y}$.
For all $y \in \mathcal{Y}$, we define

$$T^T y(\omega) \triangleq \langle y, \theta(\omega)\rangle_{\mathcal{Y},\mathcal{X}}, \omega \in \Omega \quad (2.5)$$

where $\theta : \Omega \to \mathcal{X}$ is a function such that for all $y \in \mathcal{Y}$, the function $T^T y : \Omega \to \mathbb{R}$ is measurable. The operator T is then defined for all $\ell \in \mathcal{M}^{\mathcal{V}}$ and $x \in \mathcal{X}$ by $T\ell = x$ if and only if for all $y \in \mathcal{Y}$, $\langle \ell, T^T y\rangle_{\mathcal{M}^{\mathcal{V}},\mathcal{V}} = \langle y, x\rangle_{\mathcal{X},\mathcal{Y}}$. Note that one must assume that

$$T^T\mathcal{Y} \subset \mathcal{V} \quad (2.6)$$

for this definition to be meaningful. This means that for any $\ell \in \mathcal{M}^{\mathcal{V}}$, $T\ell \in \mathcal{X}$ is characterized by:

$$\langle T\ell, y\rangle = \int_\Omega \langle y, \theta(\omega)\rangle \, Q_\ell(d\omega), \forall y \in \mathcal{Y}, \tag{2.7}$$

where $\theta : \Omega \to \mathcal{X}$ is a function as in (2.5) and $Q_\ell \in \mathcal{M}(\Omega)$ is such that $\langle \ell, v\rangle = \int_\Omega v \, dQ_\ell$ for all $v \in \mathcal{V}$.

Some examples of constraints. If one considers *moment constraints* such as $TQ = (\int_\Omega \theta_k \, dQ)_{1\le k\le K} \in \mathbb{R}^K$, $Q \in \mathcal{M}(\Omega)$ then $\theta = (\theta_k)_{1\le k\le K}$ is a measurable function from Ω to $\mathcal{X} = \mathbb{R}^K$ and $\mathcal{Y} = \mathbb{R}^K$. For this reason, one may call θ a *moment function* even for another kind of constraints.
Another example is the *marginal constraints*: Let $\Omega = \Omega_a \times \Omega_b$ be a product space and denote $Q_a = Q(\cdot \times \Omega_b)$ and $Q_b = Q(\Omega_a \times \cdot)$ the marginal measures of the probability measure Q on Ω. The constraint of prescribed marginal measures corresponds to $TQ = (Q_a, Q_b)$, $\theta(\omega_a, \omega_b) = (\delta_{\omega_a}, \delta_{\omega_b}) \in \mathcal{X} = \mathcal{P}(\Omega_a) \times \mathcal{P}(\Omega_b)$ and one can choose $\mathcal{Y} = B(\Omega_a) \times B(\Omega_b)$ where $B(\Omega)$ denotes the space of all bounded measurable functions on Ω.

2.3. **Empirical measures with random weights.** Let us consider a triangular array $\{(\omega_i^n)_{1\le i\le n}, n \ge 1\}$ of elements ω_i^n in a measurable space Ω such that the empirical measure $R_n \triangleq \frac{1}{n}\sum_{i=1}^n \delta_{\omega_i^n}$ tends to some probability measure R as n tends to infinity (in a weak sense to be made precise at (2.9)). Although the ω_i^n's may be the random outcomes of an R-iid sequence, for the present purpose, one should better think of R_n as a deterministic object. A random weight W_i^n is attached to each $\omega_i^n \in \Omega : \{W_i^n; i \le n, n \ge 1\}$ is a family of *independent* real-valued random variables. In addition, the law of W_i^n is assumed to depend on ω_i^n. We denote $W_{\omega_i^n}$ a copy of $W_i^n : \mathcal{L}aw(W_i^n) = \mathcal{L}aw(W_{\omega_i^n})$, where $\{W_\omega; \omega \in \Omega\}$ is a collection of independent random variables.
We are going to consider the large deviations as n tends to infinity of the sequence of random signed measures on Ω :

$$L_n \triangleq \frac{1}{n}\sum_{i=1}^n W_i^n \delta_{\omega_i^n} \in \mathcal{M}(\Omega), n \ge 1. \tag{2.8}$$

Let $\mathcal{R}$ be a vector subspace of $\mathcal{L}_1(\Omega, R)$: the space of R-integrable functions on Ω (where R-almost everywhere equal functions are *not* identified). We assume that R_n converges to R in the following weak sense

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n f(\omega_i^n) = \int_\Omega f \, dR, \forall f \in \mathcal{R} \subset \mathcal{L}_1(\Omega, R) \tag{2.9}$$

As a consequence of assumption (2.11) below, the mean profile

$$\omega \in \Omega \mapsto m(\omega) \triangleq \mathbb{E}W_\omega \in \mathbb{R} \tag{2.10}$$

is well defined. It is related to the law of large numbers. Indeed, it follows from the LDP of Theorem 2.16 below together with Borel-Cantelli's lemma that under our assumptions L_n converges to the signed measure mR (whose Radon-Nikodym derivative with respect to R is m) in the following sense: for all $f \in \mathcal{R}$, $\lim_{n\to\infty} \int_\Omega f \, dL_n = \int_\Omega fm \, dR$, almost surely.

Let us assume that

$$\forall \omega \in \Omega, \forall s \in \mathbb{R}, \mathbb{E}e^{sW_\omega} < \infty \tag{2.11}$$

and define the log-Laplace transform of W_ω :

$$\lambda(\omega, s) \triangleq \log \mathbb{E}e^{sW_\omega}, \omega \in \Omega, s \in \mathbb{R}.$$

The expectation $\mathbb{E}$ refers to the integration with respect to the probability measure $\mathbb{P}$ (on an unspecified space) which governs the randomness of the independent random variables $(W_\omega)_{\omega\in\Omega}$.

Let $\mathcal{V}$ be a vector space of measurable functions $v : \Omega \to \mathbb{R}$ which satisfies the following condition

$$\forall v \in \mathcal{V}, \lambda \circ v \in \mathcal{R} \subset \mathcal{L}_1(\Omega, R) \tag{2.12}$$

where $\lambda \circ v$ is the function $\omega \in \Omega \mapsto \lambda(\omega, v(\omega)) \in \mathbb{R}$. Note that this is an integrability requirement on $\mathcal{V}$: it implies that $\mathcal{V}$ is included in some Orlicz space (with respect to the measure R), see (2.17) below. Let $\mathcal{M}^{\mathcal{V}}$ be the space of linear forms on $\mathcal{V}$ which act on $\mathcal{V}$ as signed measures, see (2.4). It is endowed with the weak topology $\sigma(\mathcal{M}^{\mathcal{V}}, \mathcal{V})$ and the related Borel σ-field.

Let us introduce the convex integral functional I_{λ^*} which is defined for all signed measure Q on Ω by

$$I_{\lambda^*}(Q) \triangleq \begin{cases} \int_\Omega \lambda^*(\omega, \frac{dQ}{dR}(\omega))\, R(d\omega) & \text{if } |Q| \ll R \\ +\infty & \text{otherwise} \end{cases} \tag{2.13}$$

where $\lambda^*(\omega, t) = \sup_{s\in\mathbb{R}}\{st - \lambda(\omega, s)\}$ for all $\omega \in \Omega$ and $t \in \mathbb{R}$, is the convex conjugate of $\lambda(\omega, \cdot)$. Note that for all ω, $\lambda^*(\omega, \cdot)$ is $[0, \infty]$-valued so that the above generalized integral is meaningful.

The following functional I is the rate function of the LDP stated in Theorem 2.16 below. It is defined for all $\ell \in \mathcal{M}^{\mathcal{V}}$ by

$$I(\ell) \triangleq \inf_{Q_\ell} I_{\lambda^*}(Q_\ell)$$

where the inf is taken over all measures Q_ℓ on Ω that integrate the functions in $\mathcal{V}$ and such that their restriction to $\mathcal{V}$ is ℓ. We use the convention $\inf \emptyset = +\infty$.

In particular, if $\mathcal{V}$ is large enough to determine the measures on Ω, $\mathcal{M}^{\mathcal{V}}$ is in one to one correspondence with the space of all measures on Ω such that $\int_\Omega v\, d|\ell| < \infty$ for all nonnegative $v \in \mathcal{V}$ and we obtain for all $\ell \in \mathcal{M}^{\mathcal{V}}$,

$$I(\ell) = I_{\lambda^*}(\ell) \tag{2.14}$$

up to some obvious identification.

We also assume that the mean profile satisfies

$$mv \in \mathcal{R}, \forall v \in \mathcal{V}. \tag{2.15}$$

This is a technical requirement, which should be implied by (2.12) in standard situations since for any $v \in \mathcal{V}$ and all real $t \neq 0$, $\omega \mapsto \frac{\lambda(\omega,tv(\omega))-\lambda(\omega,0)}{t}$ belongs to the vector space $\mathcal{R}$ and its limit as t tends to 0 is mv. Note also that (2.15) is necessary to state the LLN: $\{L_n\}$ converges in law to mR in $\mathcal{M}^{\mathcal{V}}$.

Theorem 2.16 (Large deviations for empirical measures with random weights)**.** *Under the assumptions (2.9), (2.11), (2.12) and (2.15), $\{L_n\}$ satisfies the LDP in $\mathcal{M}^{\mathcal{V}}$ for the topology $\sigma(\mathcal{M}^{\mathcal{V}}, \mathcal{V})$ with the rate function I.*

Proof. For similar results see (Ellis, Gough and Puli, [11]), (Gamboa and Gassiat, [14]), (Cattiaux and Gamboa, [5]) and (Dembo and Zeitouni, [9], Theorem 7.2.3). Let us begin with a simplifying remark.
Centering. Considering the centered measure $L_n - \frac{1}{n}\sum_{i=1}^n m(\omega_i^n)\delta_{\omega_i^n}$ instead of L_n, without loss of generality one may assume that $m \equiv 0$. This corresponds to the substitution of $\lambda(\omega, s)$ by $\gamma(\omega, s) = \lambda(\omega, s) - m(\omega)s$ which is a nonnegative convex function which vanishes at 0. In particular, $\gamma_o(\omega, s) \overset{\triangle}{=} \max(\gamma(\omega, s), \gamma(\omega, -s))$ (see (3.2)) is a Young function and one can consider the Orlicz space $L_{\gamma_o^*}(\Omega, R)$ which is the topological dual space of the Orlicz space $M_{\gamma_o}(\Omega, R)$. For the definitions, notations and basic results about Orlicz spaces, see Appendix A. Our assumption (2.12) implies that

$$\mathcal{V} \subset M_{\gamma_o}(\Omega, R). \tag{2.17}$$

Without loss of generality, it is supposed during the proof that the W_ω's are centered, so that $\lambda = \gamma$. As the W_i^n's are independent random variables, for any $v \in \mathcal{V}$, we have $\Phi_n(v) \overset{\triangle}{=} \frac{1}{n}\log \mathbb{E}\exp n\langle v, L_n\rangle = \frac{1}{n}\sum_{i=1}^n \gamma(\omega_i^n, v(\omega_i^n))$. Because of assumptions (2.9) and (2.12), the limiting log-Laplace $\Phi(v) \overset{\triangle}{=} \lim_n \Phi_n(v)$ exists for every v in $\mathcal{V}$ and we have

$$\Phi(v) = \int_\Omega \gamma(\omega, v(\omega))\, R(d\omega) \in [0, \infty), v \in \mathcal{V}.$$

By assumption (2.12), Φ is finite everywhere on $\mathcal{V}$. Hence, it is a steep function. In particular, for all integer $d \geq 1$, $v_1, \dots, v_d$ in $\mathcal{V}$, the $\mathbb{R}^d$-valued random vector: $(\langle v_1, L_n\rangle, \dots, \langle v_d, L_n\rangle)$ admits the limiting log-Laplace transform $\Phi_{v_1,\dots,v_d}(\lambda_1, \dots, \lambda_n) = \Phi(\sum_{j=1}^d \lambda_j v_j)$ which is also a steep function on $\mathbb{R}^d$. Thanks to the exponential Markov inequality, for all $v \in \mathcal{V}$, there exists $\lambda > 0$ such that for all sufficiently large $A \geq 0$, $\limsup_n \frac{1}{n}\log \mathbb{P}(\langle v, L_n\rangle \geq A) \leq -\lambda A + \Phi(\lambda v)$. A standard argument allows us to infer that the finite dimensional random vector $(\langle v_1, L_n\rangle, \dots, \langle v_d, L_n\rangle)$ is exponentially tight. Applying Gärtner-Ellis theorem together with Dawson-Gärtner theorem on projective limits of LDP's, one obtains the LDP for $\{L_n\}$ on the algebraic dual space $\mathcal{V}^\sharp$ with the topology $\sigma(\mathcal{V}^\sharp, \mathcal{V})$ and the rate function Φ^*.
By assumption (2.11), under the centering convention: $m = 0$, Φ fulfils the assumption (H_Φ) of Appendix B and thanks to ([22], Lemma 6.2-b), dom Φ^* is included in the Orlicz space $L_{\gamma_o^*}(\Omega, R)$ which is the topological dual space of the Orlicz space $M_{\gamma_o}(\Omega, R)$. Define $I_\gamma(f) = \int_\Omega \gamma(\omega, f(\omega))\, R(d\omega)$ for $f \in L_{\gamma_o}(\Omega, R)$ and similarly $I_{\gamma^*}(g) = \int_\Omega \gamma^*(\omega, g(\omega))\, R(d\omega)$ for $g \in L_{\gamma_o^*}(\Omega, R)$. One can apply (Rockafellar, [28], Thm 2) to obtain that I_{γ^*} is the convex conjugate of I_γ for the duality $(M_{\gamma_o}, L_{\gamma_o^*})$. But, Φ is the restriction of I_γ to $\mathcal{V}$. By the little dual equality (B.3) applied with $T : \mathcal{V} \to M_{\gamma_o}$ being the canonical inclusion $Tv = v$, $\Phi = I_\gamma$ and $\Gamma = \Phi$: the restriction of I_γ to $\mathcal{V}$, one gets $\Phi^*(\ell) = \inf\{I_{\gamma^*}(\frac{d\tilde{\ell}}{dR}); \tilde{\ell}, \tilde{\ell}_{|\mathcal{V}} = \ell\}$. We have just shown that

$$I(\ell) = \Phi^*(\ell), \forall \ell \in \mathcal{M}^{\mathcal{V}} \text{ and dom } I = \text{dom } \Phi^*. \tag{2.18}$$

This completes the proof of the theorem. $\square$

A similar LDP with weaker integrability assumptions but stronger regularity requirements is obtained by J. Najim in [24]. For instance, [24]'s LDP holds true for $\{L_n\}$ with iid W_i^n's such that their common log-Laplace transform λ has a domain with a nonempty interior. The price to pay is an extended rate function on an extended state space possibly larger than $\mathcal{M}(\Omega)$.

With Theorem 2.16 in hand, we are going to investigate the asymptotic behavior of L_n conditionally on some event $TL_n \in C_\delta$ where the C_δ's satisfy the assumptions stated at the beginning of this section. The weak convergence in the set $\mathcal{P}(\mathcal{M}^{\mathcal{V}})$ of probability measures on $\mathcal{M}^{\mathcal{V}}$ is related to the usual topology $\sigma(\mathcal{P}(\mathcal{M}^{\mathcal{V}}), C_b(\mathcal{M}^{\mathcal{V}}))$. Loosely speaking, the following theorem states that "conditionally on $TL_n \in C_\delta$, the random measure L_n converges in law to the deterministic measure $\bar{\ell}$ which is the unique solution of (2.21)".

Let $T : \mathcal{M}^{\mathcal{V}} \to \mathcal{X}$ be a linear operator and let $\mathcal{X}$ be in separating duality with some vector space $\mathcal{Y}$. This allows to define $T^T : \mathcal{Y} \to (\mathcal{M}^{\mathcal{V}})^\sharp$ the adjoint operator of T, where $(\mathcal{M}^{\mathcal{V}})^\sharp$ is the algebraic dual space of $\mathcal{M}^{\mathcal{V}}$.

Theorem 2.19 (CLLN for empirical measures with random weights)**.** *Let us assume that the hypotheses of Theorem 2.16 are satisfied and that T satisfies the following regularity condition:*

$$T^T\mathcal{Y} \subset \mathcal{V} \tag{2.20}$$

Let $\{C_\delta\}$ be a family of subsets of $\mathcal{X}$ which fulfills the assumptions (b), (c) and (d) on the conditioning events at the beginning of this section with C a $\sigma(\mathcal{X},\mathcal{Y})$-closed convex subset of $\mathcal{X}$ which is endowed with the topology $\sigma(\mathcal{X},\mathcal{Y})$. Then,

$$\lim_{\delta\to 0}\lim_{n\to\infty} \mathbb{P}(L_n \in \cdot \mid TL_n \in C_\delta) = \delta_{\bar{\ell}} \text{ weakly in } \mathcal{P}(\mathcal{M}^{\mathcal{V}})$$

where $\bar{\ell}$ is the unique solution to the convex minimization problem

$$\text{minimize } I(\ell) \text{ subject to } T\ell \in C, \ell \in \mathcal{M}^{\mathcal{V}}. \tag{2.21}$$

Similarly, we have

$$\lim_{\delta\to 0}\lim_{n\to\infty} \mathbb{P}(X_n \in \cdot \mid X_n \in C_\delta) = \delta_{\bar{x}} \text{ weakly in } \mathcal{P}(\mathcal{X})$$

where $\bar{x}$ is the unique solution to the convex minimization problem

$$\text{minimize } J(x) \text{ subject to } x \in C \subset \mathcal{X}.$$

In the special case where T is specified by a moment function θ as in (2.5), the conditioning events $TL_n \in C_\delta$ are $\frac{1}{n}\sum_{i=1}^n W_i^n \theta(\omega_i^n) \in C_\delta$ and the assumption (2.20) is

$$\langle y, \theta(\cdot)\rangle \in \mathcal{V}, \forall y \in \mathcal{Y} \tag{2.22}$$

Note that (2.20) is the consistency requirement (2.6).

Proof. This follows immediately from Theorems 2.3 and 2.16, since the operator T is continuous, I is $\sigma(\mathcal{M}^{\mathcal{V}},\mathcal{V})$-inf-compact and strictly convex. Indeed, because of the assumption $T^T\mathcal{Y} \subset \mathcal{V}$, T is continuous with respect to the topologies $\sigma(\mathcal{M}^{\mathcal{V}},\mathcal{V})$ and $\sigma(\mathcal{X},\mathcal{Y})$. In addition, by (2.18) $I = \Phi^*$ and $\mathrm{dom}\, I = \mathrm{dom}\, \Phi^*$. But $\mathcal{V} \subset M_{\gamma_o}$ by (2.17) and $\mathrm{dom}\, \Phi^* \subset L_{\gamma_o^*}$. It follows by ([18], Cor. 2.2) that Φ^* is $\sigma(L_{\gamma_o^*}, M_{\gamma_o})$-inf-compact, so that I is $\sigma(\mathcal{M}^{\mathcal{V}},\mathcal{V})$-inf-compact . It is strictly convex since $\lambda^*(\omega,\cdot)$ is strictly convex as it is the convex conjugate of a log-Laplace transform, and a Laplace transform is Gâteaux-differentiable. $\square$

The minimizer $\bar{\ell}$ is the *entropic projection* of mR on the set of all measures ℓ such that $T\ell \in C$. This terminology is borrowed from Csiszár [6], [8], see Definition 4.2 below.

3. Good constraints

Because of Theorem 2.19, while deriving a CLLN for empirical measures with random weights with conditioning events: $\{TL_n \in C_\delta\}$, one has to minimize an entropy functional (the rate function of the underlying LDP) under the constraint $T\ell \in C$. Our aim in this section is to apply some results of the author [22] to the minimization of an entropy functional under the convex constraint: $T\ell \in C$. The relevant results of [22] are recalled below in Appendix B.

3.1. **An equivalent minimization problem.** For the moment, we are concerned with the minimization problem (2.21): the framework is Section 2.3's one with constraints defined by means of a moment function $\theta : \Omega \to \mathcal{X}$ as in (2.5) where the vector space $\mathcal{X}$ is in dual pairing with some vector space $\mathcal{Y}$ and θ is such that for all $y \in \mathcal{Y}$, the function $\omega \in \Omega \mapsto \langle y, \theta(\omega)\rangle_{\mathcal{Y},\mathcal{X}} \in \mathbb{R}$ is measurable.
Let us first *assume that $\mathcal{V}$ separates the measures* on Ω so that by (2.14), I is the λ^*-entropy I_{λ^*} which is defined at (2.13). Therefore, problem (2.21) is

$$\textit{minimize } I(Q) \textit{ subject to } \int_\Omega \theta\, dQ \in C, Q \in \mathcal{M}^{\mathcal{V}}. \tag{3.1}$$

with

$$I(Q) = I_{\lambda^*}(Q) = \begin{cases} \int_\Omega \lambda^*(\omega, \frac{dQ}{dR}(\omega))\, R(d\omega) & \text{if } |Q| \ll R \\ +\infty & \text{otherwise} \end{cases}$$

In order to state an equivalent minimization problem below at (3.4), one has to introduce some relevant Orlicz spaces. Basic definitions and results about Orlicz spaces are recalled at Appendix A. Let us consider the function

$$\gamma(\omega, s) = \lambda(\omega, s) - m(\omega)s$$

where $m(\omega)$ is the mean profile (2.10). For all ω, $s \mapsto \gamma(\omega, s)$ is a nonnegative convex function and it vanishes at 0. In particular,

$$\gamma_o(\omega, s) \triangleq \max(\gamma(\omega, s), \gamma(\omega, -s)) \tag{3.2}$$

is a Young function and one can consider the Orlicz spaces

$$\begin{aligned} M_{\gamma_o}(\Omega, R) &= \{u : \Omega \to \mathbb{R}; \forall \alpha > 0, \int_\Omega \gamma_o(\omega, \alpha u(\omega))\, R(d\omega) < \infty\} \text{ and} \\ L_{\gamma_o^*}(\Omega, R) &= \{f : \Omega \to \mathbb{R}; \exists \alpha > 0, \int_\Omega \gamma_o^*(\omega, \alpha f(\omega))\, R(d\omega) < \infty\} \end{aligned}$$

where $\gamma_o^*(\omega, \cdot)$ is the convex conjugate of $\gamma_o(\omega, \cdot)$. We identify the space of R-absolutely continuous signed measures having a density in the Orlicz space $L_{\gamma_o^*}(\Omega, R)$ with this Orlicz space. We make use of the shortcuts M_{γ_o} and $L_{\gamma_o^*}$.
We assume that $\mathcal{V}$ is a subspace of M_{γ_o}.
Beware, contrary to M_{γ_o}, $\mathcal{V}$ is not a set of R-a.e. equivalence classes. Our assumption is equivalent to $\int_\Omega \gamma(\omega, v(\omega))\, R(d\omega) < \infty$ for all $v \in \mathcal{V}$. Thanks to assumption (2.22), this implies that

$$\forall y \in \mathcal{Y}, \int_\Omega \gamma(\omega, \langle y, \theta(\omega)\rangle)\, R(d\omega) < \infty \tag{3.3}$$

As $\mathcal{V} \subset M_{\gamma_o}$, by Hölder's inequality we obtain: $L_{\gamma_o^*} \subset \mathcal{M}^{\mathcal{V}}$. Since in addition the effective domain of I is a subset of $L_{\gamma_o^*}$, we see that problem (3.1) is equivalent to

$$\textit{minimize } I(Q) \textit{ subject to } \int_\Omega \theta\, dQ \in C, Q \in L_{\gamma_o^*} \tag{3.4}$$

under our assumptions which are: $\mathcal{V}$ separates $\mathcal{M}(\Omega)$ and $\langle \mathcal{Y}, \theta \rangle \subset \mathcal{V} \subset M_{\gamma_o}$.

3.2. **The assumptions.** From now on, we drop the setting of the Empirical Measures with Random Weights. This means that we consider the minimization problem (3.4) without any spaces $\mathcal{V}$ or $\mathcal{R}$ and the function λ is not supposed to be a log-Laplace transform as in Section 2.3.

Let $(\Omega, \mathcal{A}, R)$ be a probability space where $\mathcal{A}$ is a σ-field which is supposed to be R-complete in order to be able to apply the results of [22]. Let us make some

Assumptions on λ^*. *We assume that $\lambda^*(\cdot, t)$ is ω-measurable for all t and that for R-almost every $\omega \in \Omega$, $\lambda^*(\omega, \cdot)$ is a closed strictly convex $[0, +\infty]$-valued function on $\mathbb{R}$ which attains a unique minimum. Let $m(\omega)$ denote this argmin. It is also assumed that for R-almost every $\omega \in \Omega$, the minimum value is $\lambda^*(\omega, m(\omega)) = 0$, that $\lambda^*(\omega, \cdot)$ is finite on a neighborhood of $m(\omega)$, and that there exists $A(\omega) > 0$ such that $\lambda^*(\omega, m(\omega) + A(\omega)) > 0$ and $\lambda^*(\omega, m(\omega) - A(\omega)) > 0$.*

As a consequence, I uniquely achieves its minimum value at mR and $I(mR) = 0$. As a closed strictly convex function, $t \mapsto \lambda^*(\omega, t)$ is the convex conjugate of a closed differentiable convex function $s \mapsto \lambda(\omega, s)$. Using the notation $\lambda'(\omega, s) := \frac{\partial}{\partial s}\lambda(\omega, s)$, one considers the function

$$\gamma(\omega, s) = \lambda(\omega, s) - m(\omega)s \text{ where } m(\omega) = \lambda'(\omega, 0).$$

For all ω, $s \mapsto \gamma(\omega, s)$ is a nonnegative convex function and it vanishes at 0.
Let us make a remark about measurability. As a convex function on $\mathbb{R}$, λ^* is continuous on the interior of its domain. Under our assumptions, λ^* is jointly measurable, and so are λ and λ'. Hence, m is a measurable function and so is γ. Let us make some

Assumptions on θ. *We assume that*

$$\begin{aligned} &\forall y \in \mathcal{Y}, \int_\Omega \gamma(\omega, \langle y, \theta(\omega) \rangle)\, R(d\omega) < \infty \quad \textit{and} \\ &\forall y \in \mathcal{Y}, \langle y, \theta(\cdot) \rangle = 0, R\text{-a.e.} \Rightarrow y = 0 \end{aligned} \tag{3.5}$$

The first assumption is (3.3). Since $\mathcal{X}$ and $\mathcal{Y}$ are in separating duality, the second requirement (3.5) states that the vector space spanned by the range of θ "is essentially" $\mathcal{X}$. This is not an effective restriction. Let us make some

Assumptions on C. *We assume that C is a $\sigma(\mathcal{X}, \mathcal{Y})$-closed convex subset of $\mathcal{X}$.*

3.3. **What good and bad constraints are.** If the Young function γ_o doesn't satisfy the Δ_2-condition (see (A.1)), for instance if it has an exponential growth at infinity, $M_{\gamma_o} = \{u : \Omega \to \mathbb{R}; \forall \alpha > 0, \int_\Omega \gamma_o(\omega, \alpha u(\omega))\, R(d\omega) < \infty\}$ may be a proper subset of

$$L_{\gamma_o} = \{u : \Omega \to \mathbb{R}; \exists \alpha > 0, \int_\Omega \gamma_o(\omega, \alpha u(\omega))\, R(d\omega) < \infty\}.$$

Consequently, for some moment functions θ, assumption (3.3) may not be satisfied while the weaker property

$$\forall y \in \mathcal{Y}, \exists \alpha > 0, \int_\Omega \gamma(\omega, \alpha\langle y, \theta(\omega)\rangle)\, R(d\omega) < \infty \tag{3.6}$$

may be shared. In this situation, analytical complications occur (see Section 4). This is the reason why constraints satisfying (3.3) are called *good constraints*, while constraints satisfying (3.6) but not (3.3) are called *bad constraints*.

3.4. **Attainment and representation of the minimizers.** Let us consider the following couple of convex conjugates

$$\begin{aligned}
\Lambda(y) &\triangleq \int_\Omega \lambda(\omega, \langle y, \theta(\omega)\rangle)\, R(d\omega), \quad y \in \mathcal{Y}\\
\Lambda^*(x) &\triangleq \sup_{y\in\mathcal{Y}} \left\{ \langle y, x\rangle - \int_\Omega \lambda(\omega, \langle y, \theta(\omega)\rangle)\, R(d\omega) \right\}, \quad x \in \mathcal{X}
\end{aligned}$$

Let us denote

$$\mathrm{dom}\ \Lambda^* = \{x \in \mathcal{X}; \Lambda^*(x) < \infty\}$$

the effective domain of Λ^* and icordom Λ^* the intrisic core of dom Λ^*. That is

$$\mathrm{icordom}\ \Lambda^* = \{x \in \mathrm{dom}\ \Lambda^*; \forall x' \in \mathrm{affdom}\ \Lambda^*, \exists t > 0, [x, x + t(x' - x)[\in \mathrm{dom}\ \Lambda^*\}$$

where affdom Λ^* is the affine space generated by dom Λ^*.
Applying Theorem B.2, we shall obtain the following

Theorem 3.7 (The minimizers of (3.4))**.** *Under the assumptions on λ^*, θ and C which are stated above (in particular the "good constraint" requirement (3.3) is assumed), we have*

$$\begin{aligned}
&\inf \left\{ I(Q); Q \in L_{\gamma_o^*}, \int_\Omega \theta\, dQ \in C \right\}\\
&= \sup_{y\in\mathcal{Y}} \left\{ \inf_{x\in C} \langle y, x\rangle - \int_\Omega \lambda(\omega, \langle y, \theta(\omega)\rangle)\, R(d\omega) \right\}\\
&= \inf_{x\in C} \Lambda^*(x) \in [0, \infty].
\end{aligned}$$

In particular, one has

$$J(x) \triangleq \inf\{I(Q); Q \in L_{\gamma_o^*}, \int_\Omega \theta\, dQ = x\} = \Lambda^*(x), x \in \mathcal{X}.$$

If $C\cap$dom Λ^* is nonempty, then (3.4) is attained and any minimizing sequence $\sigma(L_{\gamma_o^*}, M_{\gamma_o})$-converges to the unique solution $\bar{Q}$ of (3.4).
Suppose that in addition, $C \cap$icordom Λ^* is nonempty, then there exists some linear form $\bar{y}$ on $\mathcal{X}$ such that $\langle \bar{y}, \theta(\cdot)\rangle$ is measurable and

(a) $\bar{x} \triangleq \int_\Omega \theta\, d\bar{Q} \in C \cap \mathrm{dom}\ \Lambda^*$
(b) $\langle \bar{y}, \bar{x}\rangle \le \langle \bar{y}, x\rangle, \forall x \in C \cap \mathrm{dom}\ \Lambda^*$
(c) $\bar{Q}(d\omega) = \lambda'(\omega, \langle \bar{y}, \theta(\omega)\rangle)\, R(d\omega)$.

In this situation, $\bar{x}$ minimizes Λ^* on C, $I(\bar{Q}) = \Lambda^*(\bar{x})$ and because of (a) and (c), we have

$$\bar{x} = \int_\Omega \theta(\omega)\lambda'(\omega, \langle \bar{y}, \theta(\omega)\rangle)\, R(d\omega) \tag{3.8}$$

in the weak sense.

Following the terminology of Ney [25], [26], as it shares the properties (a), (b) and (3.8), the minimizer $\bar{x}$ is called a *dominating point* of C for the rate function Λ^* (see Definition 6.2 below).
Note that $\bar{y}$ does not necessarily belong to $\mathcal{Y}$. Therefore, the Young equality $\langle \bar{x}, \bar{y}\rangle = \Lambda^*(\bar{x}) + \Lambda(\bar{y})$ is meaningless. Nevertheless, there exists a natural extension $\overline{\Lambda}$ of Λ such that $\langle \bar{x}, \bar{y}\rangle = \Lambda^*(\bar{x}) + \overline{\Lambda}(\bar{y})$ holds, see (B.4) at Appendix B.

Proof. During this proof, the notations and definitions of Appendix B will be used.

Let us first establish the connection with the framework of Appendix B. We introduce the convex functional

$$\Phi(u) \triangleq \int_\Omega \gamma(\omega, u(\omega))\, R(d\omega), u \in \mathcal{U}_0$$

where $\mathcal{U}_0 = M_{\gamma_o}$. We have

$$I(Q) = \Phi^*(Q - mR) \tag{3.9}$$

where Φ^* is the convex conjugate of Φ for the duality $(\mathcal{L}_0, \mathcal{U}_0)$ with $\mathcal{L}_0 \triangleq \mathcal{U}_0^\sharp$: the algebraic dual space of $\mathcal{U}_0$. That is

$$\Phi^*(\ell) = \sup_{u \in \mathcal{U}_0} \{\langle \ell, u\rangle - \Phi(u)\} \in [0, \infty], \quad \ell \in \mathcal{L}_0.$$

The identity (3.9) is a consequence of general results of Rockafellar on conjugate duality for integral functionals [28]. More precisely, it is also proved in ([18], Proposition 6.2) that $\Phi^*(\ell) = \int_\Omega \gamma^*(\omega, \frac{d\ell}{dR}(\omega))\, R(d\omega)$ if ℓ is an R-absolutely continuous measure, and $\Phi^*(\ell) = \infty$ otherwise. This implies that up to the transformation $\ell = Q - mR \in \mathcal{L}_0$, the minimization problem (3.4) becomes

$$\textit{minimize } \Phi^*(\ell) \textit{ subject to } \langle \theta, \ell\rangle \in C_o, \ \ell \in \mathcal{L}_0$$

where $C_o = C - \langle \theta, mR\rangle \subset \mathcal{X}$, ($\langle \theta, \ell\rangle$ is defined in the weak sense).
The hypothesis (H_Φ) of Theorem B.2 is clearly satisfied under our assumptions on λ^*. The hypothesis (H_{T1}) is (3.3) while (H_{T2}) is (3.5).

Now, a direct application of Theorem B.2 with $\Phi(u) = \int_\Omega \gamma(u)\, dR$, $|u|_\Phi = \|u\|_{\gamma_o}$, $u \in \mathcal{U}_0 = \mathcal{U}_1 = M_{\gamma_o}$, $\mathcal{L}_1 = L_{\gamma_o^*}$ and $\Gamma(y) \triangleq \int_\Omega \gamma(\omega, \langle y, \theta(\omega)\rangle)\, R(d\omega) \in [0, \infty]$, $y \in \mathcal{Y}$ gives us the dual equality and the statement about the minimizing sequence.
The uneasy remaining work is the computation of $\bar{\Phi}$. But this has been performed by the author in ([18], Theorem 6.3). Being careless with annoying details, the result is essentially

$$\bar{\Phi}(u) = \int_\Omega \gamma(u^a)\, dR + \sup\{\langle u^s, f\rangle; f \in \text{dom } \Phi^*\}, u \in \text{dom } \bar{\Phi} \subset \mathcal{U}_2$$

where $u = u^a + u^s$ is a unique decomposition of u into a measurable function u^a and a singular part u^s. This singular part will not play any role in the rest of the proof since one only needs to compute the subdifferential $\partial\bar\Phi(u)$ and

$$\partial\bar\Phi(u) = \{\gamma'(u^a).R\} \tag{3.10}$$

for any u in dom $\bar\Phi$. To make things easier, we have supposed that λ^* is strictly convex, so that γ is differentiable. To prove (3.10), note that for any $h \in \mathcal{U}_0 = M_{\gamma_o}$ and any $u \in$ dom $\bar\Phi$, we have $(u+h)^a = u^a + h$ and $(u+h)^s = u^s$ since $h^s = 0$. Hence, $\bar\Phi(u+h) - \bar\Phi(u) = [\int_\Omega \gamma(u^a+h)\,dR + \sup\{\langle u^s, f\rangle; f \in \text{dom } \Phi^*\}] - [\int_\Omega \gamma(u^a)\,dR + \sup\{\langle u^s, f\rangle; f \in \text{dom } \Phi^*\}] = \int_\Omega \gamma(u^a+h)\,dR - \int_\Omega \gamma(u^a)\,dR$.
As the transformation $\lambda \to \gamma$ corresponds to the transformation $Q \to \ell = Q - mR$, the minimizer of (3.4) is $\bar Q = \lambda'((T^*y_*)^a)\cdot R$ for some $y_* \in \mathcal{Y}_2$. ☐

4. Bad constraints

In this section, the minimization problem (3.4) is considered when the constraint function θ satisfies (3.6) but not necessarily (3.3). This means that the constraint is *bad.* Note that this situation is of interest if M_{γ_o} is a proper subspace of L_{γ_o} since (3.3) is equivalent to $\langle\mathcal{Y},\theta\rangle \subset M_{\gamma_o}$ and (3.6) is equivalent to $\langle\mathcal{Y},\theta\rangle \subset L_{\gamma_o}$. This occurs if and only if γ_o doesn't share the Δ_2 property, see (A.1). On the other hand, thanks to Hölder's inequality in Orlicz spaces, the constraint integrals $\int_\Omega \langle y,\theta\rangle\,dQ$, $y \in \mathcal{Y}$ are well defined for all $Q \in L_{\gamma_o^*}$ whenever $\langle\mathcal{Y},\theta\rangle \subset L_{\gamma_o}$.

An important instance of this non-Δ_2 situation is encountered with the *relative entropy* $I(Q \mid R)$ of the probability measure Q with respect to the probability measure R. This corresponds to $\lambda^*(t) = t\log t - t + 1$ adding the mass constraint $\int_\Omega \mathbf{1}\,dQ = 1$. Note that the nonnegativity of Q is ensured by dom $\lambda^* = [0,\infty)$. With this special choice of λ^*, one obtains $\lambda(s) = e^s - 1$ and $\gamma_o(s) = e^{|s|} - |s| - 1$ which doesn't share the Δ_2 property. This special situation is developed in Section 5 below.

Under these assumptions, problem (3.4) may not be attained anymore. As will soon be seen, this phenomenon is tightly linked to the notion of *generalized entropic projection* introduced by Csiszár.

4.1. Generalized entropic projections.

In [8], Csiszár has proved the following result.

Theorem 4.1 (Csiszár)**.** *Whenever* dom $\lambda^* \subset [0,\infty)$ *and* $\lambda^{*\prime}(\infty) = \infty$, *for any* $\mathcal{C}$ *convex subset of the set* $\mathcal{M}_+(\Omega)$ *of all nonnegative measures on* Ω *such that* $\mathcal{C}\cap$dom I *is nonempty, any minimizing sequence of*

$$\text{minimize } I(Q) \text{ subject to } Q \in \mathcal{C}, Q \in \mathcal{M}_+(\Omega)$$

converges in variation norm to some $Q_* \in \mathcal{M}_+(\Omega)$.

Definition 4.2 (Csiszár)**.** *This* Q_* *which is called the* generalized entropic projection *of* mR *on* $\mathcal{C}$ *(with respect to* I*) may not belong to* $\mathcal{C}$. *In case* Q_* *is in* $\mathcal{C}$, *it is called the* entropic projection *of* mR *on* $\mathcal{C}$.

Let us first show that as a direct consequence of Theorems 3.7 and 4.1, if the constraints are good, the generalized entropic projection is the entropic projection. We are concerned with the special case where $\mathcal{C} = \{Q \in \mathcal{M}(\Omega) \cap \text{dom } T; TQ := \int_\Omega \theta\,dQ \in C\}$ which leads us to the minimization problem (3.4):

$$\textit{minimize } I(Q) \textit{ subject to } \int_\Omega \theta\,dQ \in C, Q \in L_{\gamma_o^*} \tag{P}$$

Proposition 4.3. *Under the assumption of Theorem 3.7, if $C \cap \operatorname{dom} \Lambda^*$ is nonempty, then the entropic projection Q_* of mR on $\mathcal{C} = \{Q \in \mathcal{M}(\Omega); \int_\Omega \theta\, dQ \in C\}$ exists. It is $\bar{Q} \in \mathcal{C}$: the minimizer of (P)=(3.4) and it is described at Theorem 3.7.*

Proof. At Theorem 3.7, it is proved that if $C \cap \operatorname{dom} \Lambda^*$ is nonempty, then (3.4) is attained and any minimizing sequence $\sigma(L_{\gamma_o^*}, M_{\gamma_o})$-converges to the unique solution $\bar{Q}$ of (3.4). Therefore, any minimizing sequence is a fortiori convergent for the topology $\sigma(L_1, L_\infty)$. By Theorem 4.1, it converges in variation to some Q_*. Hence, $Q_* = \bar{Q}$. $\square$

Csiszár's proof of Theorem 4.1 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_*. Our purpose in the present section is to bring details on the generalized entropic projection in specific situations. To do so, let us introduce new objects. We shall take advantage of the study of the following extended minimization problem

$$\textit{minimize } \bar{I}(\ell) \textit{ subject to } \langle \theta, \ell \rangle \in C, \ell \in L'_{\gamma_o} \qquad (\bar{P})$$

where L'_{γ_o} is the topological dual space of the Orlicz space L_{γ_o} and $\bar{I}$ is the greatest convex $\sigma(L'_{\gamma_o}, L_{\gamma_o})$-lower semi-continuous extension of I to $L'_{\gamma_o} \supset L_{\gamma_o^*}$. The dual space L'_{γ_o} admits the representation $L'_{\gamma_o} \simeq L_{\gamma_o^*} \oplus L^s_{\gamma_o}$. This means that any $\ell \in L'_{\gamma_o}$ is uniquely decomposed as $\ell = \ell^a + \ell^s$ where $\ell^a \in L_{\gamma_o^*}$ and $\ell^s \in L^s_{\gamma_o}$ are respectively the *absolutely continuous part* and the *singular part* of ℓ. For more details, see Appendix A. The extension $\bar{I}$ has the following form, see (Fougères and Giner, [12], Thm. 3.2),

$$\bar{I}(\ell) = I(\ell^a) + I^s(\ell^s), \ell \in L'_{\gamma_o} \qquad (4.4)$$

where $\ell = \ell^a + \ell^s$ is the above decomposition of ℓ and $I^s(\ell^s) \geq 0$ is the recession function of $\bar{I}$:

$$I^s(\ell^s) = \sup \left\{ \langle \ell^s, u \rangle; u \in L_{\gamma_o}, \int_\Omega \gamma(u)\, dR < \infty \right\} \in [0, \infty].$$

For all $\ell \in L'_{\gamma_o}$ and $x \in \mathcal{X}$, we define

$$T\ell = \langle \theta, \ell \rangle = x \text{ if and only if } \Big\langle \langle y, \theta \rangle, \ell \Big\rangle_{L_{\gamma_o}, L'_{\gamma_o}}, \forall y \in \mathcal{Y}.$$

Bad constraints assumptions. *Let us assume that*

- *λ^* satisfies the same assumptions as in Section 3,*
- *θ satisfies (3.5), but the "bad constraint" assumption (3.6) instead of the "good constraint" assumption (3.3), and*
- *C is a $\sigma(\mathcal{X}, \mathcal{Y})$-closed convex subset of $\mathcal{X}$.*

Theorem 4.5. *Suppose that the bad constraints assumptions hold. Then, the dual equality*

$$\inf\{\bar{I}(\ell); \langle \theta, \ell \rangle \in C, \ell \in L'_{\gamma_o}\} = \sup_{y \in \mathcal{Y}} \left\{ \inf_{x \in C} \langle y, x \rangle - \int_\Omega \lambda(\langle y, \theta \rangle)\, dR \right\} \in [0, \infty] \qquad (4.6)$$

holds.
Suppose that in addition $C \cap \operatorname{dom} \Lambda^*$ is nonempty. Then, the minimization problem $(\bar{P})$ is attained in L'_{γ_o} and all its solutions share the same unique absolutely continuous part $Q_* \in L_{\gamma_o^*}$. This means that for all ℓ_* solution to $(\bar{P})$, we have $\ell_*^a = Q_*$.

Proof. This proof relies upon results which are stated and proved in the remainder of the present section.

The dual equality (4.6) is a direct consequence of part (b) of Lemma 4.15 and part (c) of Lemma 4.16.

The attainment in $(\bar{P})$ and the statement about the absolutely continuous part of the minimizers are proved in parts (b) and (d) of Lemma 4.9. $\square$

Theorem 4.7. *Suppose that the bad constraints assumptions hold. Let us assume in addition that there are finitely many constraints, i.e. $\mathcal{X}$ is finite dimensional, and $C \cap$ icordom Λ^* is nonempty. Then,*

$$\inf(P) = \inf(\bar{P}) \tag{4.8}$$

and any minimizing sequence (Q_n) of (P) converges in the sense of the $\sigma(L_{\gamma_o^*}, M_{\gamma_o})$-topology to Q_*.
Therefore, if dom $\lambda^* \subset [0,\infty)$ *and* $\lambda^{*\prime}(\infty) = \infty$ *as in Theorem 4.1, Q_* is the generalized entropic projection of mR on $\mathcal{C} = \{Q \in \mathcal{M}_+(\Omega); \int_\Omega \theta\, dQ \in C\}$.*

Proof. This proof relies upon results which are stated and proved in the remainder of the present section.

It is shown at Proposition 4.10 that any minimizing sequence (Q_n) of (P) converges in the sense of the $\sigma(L_{\gamma_o^*}, M_{\gamma_o})$-topology to Q_*, whenever $\inf(P) = \inf(\bar{P})$.
By part (a) of Lemma 4.15, for this equality to hold, it is enough that some value function φ_M (see (4.12)) is lower semi-continuous at 0. This is true under our assumptions by part (a) of Lemma 4.16. $\square$

Theorem 4.7 extends Csiszár's result ([8], Theorem 4).

As a consequence of Theorem 4.7, for any $\ell_* = \ell_*^a + \ell_*^s = Q_* + \ell_*^s$ minimizer of $(\bar{P})$ and any (Q_n) minimizing sequence of (P), one obtains

$$\begin{aligned} \inf_n I(Q_n) &= \inf(P) = \inf(\bar{P}) = \bar{I}(\ell_*) \\ &= I(Q_*) + I^s(\ell_*^s) \\ &\geq I(Q_*) \end{aligned}$$

with a strict inequality if $I^s(\ell_*^s) > 0$. This last quantity is precisely the *gap* of lower $\sigma(M', M)$-semicontinuity of I : $\lim_n Q_n = Q_*$ and $\inf(P) = \liminf_n I(Q_n) \geq I(\lim_n Q_n) = I(Q_*)$.

4.2. **Preliminary results.** Preliminary results for the proof of Theorem 4.5 are stated below at Lemma 4.9 and Proposition 4.10.
Notations. From now on, we shall denote $M = M_{\gamma_o}$ and $L = L_{\gamma_o}$, so that $M \subset L$, $L_{\gamma_o^*} = M'$ and $L' \simeq M' \oplus L^s$.

Lemma 4.9.

(a) *Under the assumption (3.6): $\langle \mathcal{Y}, \theta \rangle \subset L$, the constraint operator $T : L' \to \mathcal{X}$ is $\sigma(L', L)$-$\sigma(\mathcal{X}, \mathcal{Y})$-continuous.*
(b) *As λ^* is strictly convex, if k_* and ℓ_* are two solutions of $(\bar{P})$, their absolutely continuous parts match: $k_*^a = \ell_*^a$.*
(c) *Under the assumptions on λ^*, $\bar{I}$ is $\sigma(L', L)$-inf-compact.*

(d) *If in addition to these assumptions, C is supposed to be $\sigma(\mathcal{X},\mathcal{Y})$-closed and* $\mathrm{dom}\,\Lambda^*\cap C$ *is nonempty, then $(\bar{P})$ is attained.*

Proof. Statement (a) is straightforward. For a proof of statement (c), see for instance ([18], Corollary 2.2) or ([22], Lemma 6.6). Statement (d) is a direct consequence of (a) and (c).
Let us prove (b). Let k_*, ℓ_* be two solutions of $(\bar{P})$. They are in the convex set $\mathcal{C} = \{\ell \in L'; T\ell \in C\}$ and $\inf(\bar{P}) = \bar{I}(k_*) = \bar{I}(\ell_*)$. For all $0 \le p, q \le 1$ such that $p+q=1$, as I and I^s are convex functions, we have

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

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 $0<p,q<1$, one gets: $I(pk_*^a+q\ell_*^a) < pI(k_*^a)+qI(\ell_*^a)$ 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 (b). $\square$

Proposition 4.10. *Under the assumptions of Theorem 4.5, suppose that the identity* $\inf(P) = \inf(\bar{P})$ *holds. Then, any minimizing sequence (Q_n) of (P) $\sigma(M', M)$-converges to $Q_* \in M'$: the absolutely continuous part shared by the solutions of $(\bar{P})$ (see Lemma 4.9).*
Under the additional assumptions of Theorem 4.1, the generalized entropic projection exists, it is Q_*.

Proof. Since it is assumed that $\inf(P) = \inf \bar{P})$ holds, (Q_n) is also a minimizing sequence of $(\bar{P})$. As $\bar{I}$ is $\sigma(L', L)$-inf-compact by Lemma 4.9, one can extract a $\sigma(L', L)$-convergent subsequence $(\tilde{Q}_n)$ from (Q_n). Let $\ell_* \in \mathcal{C}$ denote its limit: we have $\lim_n \int_\Omega u\,d\tilde{Q}_n = \langle \ell_*, u\rangle$ for all $u \in L$. As $\langle \ell_*^s, u\rangle = 0$, for all $u \in M$ (see Appendix A), we obtain: $\lim_n \int_\Omega u\,d\tilde{Q}_n = \int_\Omega u\,d\ell_*^a$ for all $u \in M$. This proves that $(\tilde{Q}_n)$ $\sigma(M', M)$-converges to ℓ_*^a. Let Q_* be defined as in Lemma 4.9. By this lemma, ℓ_* is a minimizer of $(\bar{P})$ and $\ell_*^a = Q_*$. Therefore, any convergent subsequence of (Q_n) converges to Q_*. As any subsequence of a minimizing sequence is still a minimizing sequence, we have proved that from any subsequence of (Q_n), one can extract a sub-subsequence which converges to Q_*. This proves that (Q_n) $\sigma(M', M)$-converges to Q_*.
One proves the last statement as Proposition 4.3. $\square$

4.3. **Sufficient conditions for** $\inf(P) = \inf(\bar{P})$**.** Our aim now is to obtain sufficient conditions for the identity (4.8): $\inf(P) = \inf(\bar{P})$ to hold. Let us rewrite the problems (P) and $(\bar{P})$ in order to emphasize their differences and analogies. As at (3.9), let us consider the convex conjugates

$$\begin{aligned}\Phi_M^*(\ell) &= \sup_{u\in M}\{\langle \ell, u\rangle - \Phi(u)\}, \ell \in M'\\ \Phi_L^*(\ell) &= \sup_{u\in L}\{\langle \ell, u\rangle - \Phi(u)\}, \ell \in L'\end{aligned}$$

of

$$\Phi(u) = I_\gamma(u) = \int_\Omega \gamma(u)\, dR, u \in L.$$

It has been proved in (Kozek, [16], Thm 2.6), that

$$\begin{aligned} I(\ell) &= \Phi^*_M(\ell - mR), \forall \ell \in M', \\ \bar{I}(\ell) &= \Phi^*_L(\ell - mR), \forall \ell \in L' \end{aligned}$$

Hence, considering the minimization problems

$$\textit{minimize } \Phi^*_M(\ell) \textit{ subject to } T\ell \in C_o, \ell \in M' \tag{P_M}$$

and

$$\textit{minimize } \Phi^*_L(\ell) \textit{ subject to } T\ell \in C_o, \ell \in L' \tag{P_L}$$

with $C_o = C - T(mR)$, we see that ℓ_* is a solution of (P) [resp. $(\bar{P})$] if and only if $\ell_* - mR$ is a solution of (P_M) [resp. (P_L)]. The analogue of $\Lambda^*(x)$ is $\Gamma^*(x) = \Lambda^*(x - T(mR)) = \sup_{y\in\mathcal{Y}}\{\langle y, x\rangle - \Gamma(y)\}$ with $\Gamma(y) = \int_\Omega \gamma(\langle y, \theta\rangle)\, dR$.
Clearly, it will be enough to prove $\inf(P_M) = \inf(P_L)$ to get $\inf(P) = \inf(\bar{P})$.

Basic facts about convex duality. Our proof will rely on usual convex duality considerations. Let us recall some basic facts about this. Consider the *primal minimization problem*

$$\textit{minimize } f(a), a \in A \tag{$\mathcal{P}$}$$

where A is a vector space and f is a $[-\infty, \infty]$-valued function on A. Let us introduce the family of perturbed problems

$$\textit{minimize } F(a, x), a \in A \tag{$\mathcal{P}_x$}$$

where x belongs to some vector space $\mathcal{X}$ and $x = 0$ corresponds to "no perturbation": $F(\cdot, 0) = f$. Let us consider the value-function $\varphi(x) \overset{\triangle}{=} \inf(\mathcal{P}_x) = \inf_a F(a, x)$, $x \in \mathcal{X}$ and the concave conjugate of $-F(a, \cdot) : K(a, y) \overset{\triangle}{=} \inf_x\{\langle y, x\rangle + F(a, x)\}$ where y stands in some vector space $\mathcal{Y}$ in separating duality with $\mathcal{X}$. Let us define the *dual maximization problem*

$$\textit{maximize } g(y), y \in \mathcal{Y} \tag{$\mathcal{D}$}$$

where $g(y) \overset{\triangle}{=} \inf_a K(a, y)$. For all $y \in \mathcal{Y}$, we have $g(y) = \inf_{a,x}\{\langle y, x\rangle + F(a, x)\} = \inf_x\{\langle y, x\rangle + \varphi(x)\}$. Hence, for all $y \in \mathcal{Y}$, we have $-g(-y) = \sup_x\{\langle y, x\rangle - \varphi(x)\} = \varphi^*(y)$ and for all $x \in \mathcal{X}$, $\varphi^{**}(x) = \sup_y\{\langle y, x\rangle + g(-y)\}$, where φ^* and φ^{**} are the dual conjugate and biconjugate of φ for the duality $(\mathcal{X}, \mathcal{Y})$. In particular, with $x = 0$, one gets $\varphi^{**}(0) = \sup(\mathcal{D})$. As $\varphi(0) = \inf(\mathcal{P})$, we have

$$\sup(\mathcal{D}) = \varphi^{**}(0) \le \varphi(0) = \inf(\mathcal{P}) \tag{4.11}$$

If the perturbation $F(a, x)$ is jointly convex in (a, x), the value function φ is also convex. As φ^{**} is the $\sigma(\mathcal{X}, \mathcal{Y})$-closed convex regularization of φ, we obtain that the *dual equality:* $\inf(\mathcal{P}) = \sup(\mathcal{D})$ holds if and only φ is $\sigma(\mathcal{X}, \mathcal{Y})$-lower semicontinuous at $x = 0$.

In the situation where one wants to minimize the function h on A subject to the constraint $Ta \in C$ where $T : A \to \mathcal{X}$ is a linear operator and C is a subset of the vector space $\mathcal{X}$, problem $\mathcal{P}$ is

$$\textit{minimize } h(a) \textit{ subject to } Ta \in C, a \in A$$

This corresponds to $f(a) = h(a) + \delta_C(Ta)$ where $\delta_C(x) = \begin{cases} 0 & \text{if } x \in C \\ +\infty & \text{if } x \notin C \end{cases}$ is the convex indicator of C. An interesting perturbation of this problem is given by $F(a, x) = h(a) + \delta_C(Ta + x)$, $a \in A$, $x \in \mathcal{X}$. If h is a convex function and C is a convex set, then F is a convex function and so is

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

Let $\mathcal{Y}$ be in separating duality with $\mathcal{X}$. The Lagrangian is $K(a, y) = \inf_x\{\langle y, x\rangle + h(a) + \delta_C(Ta + x)\} = \inf_{x\in C}\langle y, x\rangle - \langle Ta, y\rangle + h(a)$ and $g(y) = \inf_{x\in C}\langle y, x\rangle + \inf_a\{-\langle a, T^T y\rangle + h(a)\}$. This leads us to

$$g(y) = \inf_{x\in C}\langle y, x\rangle - h^*(T^T y)$$

where $T^T\mathcal{Y}$ is a subspace of the algebraic dual space of A and h^* is the convex conjugate of h for the duality $(A, T^T\mathcal{Y}) : h^*(T^T y) = \sup_a\{\langle y, Ta\rangle - h(a)\}$.

Back to our problem. Now, let us particularize this framework for the problems (P_M) and (P_L). Assuming that $m \equiv 0$, one sees that $(P) = (P_M)$, $(\bar{P}) = (P_L)$, $I = \Phi_M^*$, $\bar{I} = \Phi_L^*$, $\lambda = \gamma$, $C = C_o$ and so on. *This simplifying requirement will be assumed during the proof, without loss of generality.* Our assumption about the bad constraint is $T^T\mathcal{Y} \subset L$.

Let us begin with (P_M). This corresponds to $A = M'$ and $h = \Phi_M^*$. Let us denote φ_M and g_M the corresponding functions φ and g. We have

$$\begin{aligned} \varphi_M(x) &= \inf\{\Phi_M^*(\ell); T\ell \in C - x, \ell \in M'\}, x \in \mathcal{X} \text{ and} \quad (4.12) \\ g_M(y) &= \inf_{x\in C}\langle y, x\rangle - I_\lambda(T^T y), y \in \mathcal{Y} \quad (4.13) \end{aligned}$$

where the last equality follows from $(\Phi_M^*)^* = I_\lambda$ where the considered duality is (M', L), see (Rockafellar, [28]). Hence, the dual problem to (P_M) is

$$\text{maximize } \inf_{x\in C}\langle y, x\rangle - I_\lambda(T^T y), y \in \mathcal{Y} \quad (D_M)$$

Let us go on with (P_L). This corresponds to $A = L'$ and $h = \Phi_L^*$. Let us denote φ_L and g_L the corresponding function φ and g. By ([21], Proposition 4.3), we obtain that the convex biconjugate of Φ_L for the duality (L', L) is also I_λ, therefore we have $g_L = g_M$ and it follows that $\sup(D_M) = \sup(D_L)$ since the dual problems (D_M) and (D_L) of (P_M) and (P_L) match. As Φ_L^* and Φ_M^* match on M', we have $\inf(P_L) \le \inf(P_M)$, so that

$$\sup(D_M) = \sup(D_L) \le \inf(P_L) \le \inf(P_M). \quad (4.14)$$

Therefore, for the desired equality $\inf(P_L) = \inf(P_M)$ to hold, it is enough that the dual equality $\inf(P_M) = \sup(D_M)$ holds. And this happens if and only if φ_M is $\sigma(\mathcal{X}, \mathcal{Y})$-lower semicontinuous at $x = 0$. We have proved the following

Lemma 4.15. *(a) If φ_M is $\sigma(\mathcal{X}, \mathcal{Y})$-lower semicontinuous at $x = 0$, then* $\inf(P_M) = \inf(P_L)$.
(b) Similarly, if φ_L is $\sigma(\mathcal{X}, \mathcal{Y})$-lower semicontinuous at $x = 0$, then $\sup(D_L) = \inf(P_L)$.

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

Lemma 4.16.

(a) *Suppose that there are finitely many constraints, i.e. $\mathcal{X}$ is finite dimensional. If* $C \cap \operatorname{icordom} \Lambda^*$ *is nonempty, then φ_M is continuous at* 0.

(b) *Suppose that (3.3) is satisfied and C is $\sigma(\mathcal{X},\mathcal{Y})$-closed, then φ_M is $\sigma(\mathcal{X},\mathcal{Y})$-lower semicontinuous.*
(c) *Suppose that (3.6) is satisfied and C is $\sigma(\mathcal{X},\mathcal{Y})$-closed, then φ_L is $\sigma(\mathcal{X},\mathcal{Y})$-lower semicontinuous.*

Proof. To get (a), simply remark that a convex function on a finite dimensional space is continuous on the interior of its effective domain. The assumption $C \cap \operatorname{icordom} \Lambda^* \neq \emptyset$ implies that 0 belongs to $\operatorname{icordom} \varphi_M$.

Let us show (b). Defining $\tilde{\varphi}(x) := \varphi(-x)$ and $J(x) := \inf\{\Phi_M^*(\ell); T\ell = x\}$, $x \in \mathcal{X}$, we obtain that $\tilde{\varphi}$ is the inf-convolution of J and the convex indicator of $-C$: δ_{-C}. That is $\tilde{\varphi}(x) = (J \Box \delta_{-C})(x) = \inf\{J(y) + \delta_{-C}(z); y, z, y + z = x\}$.
The function Φ_M^* is $\sigma(M', M)$-inf-compact and under the assumption (3.3), T is $\sigma(M', M)$-$\sigma(\mathcal{X},\mathcal{Y})$-continuous. It follows that J is $\sigma(\mathcal{X},\mathcal{Y})$-inf-compact. As C is assumed to be $\sigma(\mathcal{X},\mathcal{Y})$-closed, δ_{-C} is lower semi-continuous. Finally, being the inf-convolution of an inf-compact function and a lower semi-continuous function, $\tilde{\varphi}$ is lower semi-continuous, and so is φ_M.

The proof of (c) is similar since Φ_L^* is $\sigma(L', L)$-inf-compact and under the assumption (3.6), T is $\sigma(L', L)$-$\sigma(\mathcal{X},\mathcal{Y})$-continuous. □

5. The special case of relative entropy

Using Poissonian random weights W_ω one is able to convert the results of the previous sections into statement about relative entropy. An almost straightforward translation of Theorem 3.4, Proposition 4.3 and Theorem 4.5 is stated at Propositions 5.3 and 5.5 below. The main result of this section: Theorem 5.9, states that in a Banach space setting, the generalized I-projection is the absolutely continuous part of the minimizers of the extended relative entropy. This result is similar to Theorem 4.7, but this time no finite dimensional constraint is assumed.

Recall that the relative entropy of the probability measure P with respect to the probability measure R is defined by

$$I(P \mid R) = \begin{cases} \int_\Omega \log\left(\frac{dP}{dR}\right) dP & \text{if } P \ll R \\ +\infty & \text{otherwise.} \end{cases}.$$

The minimization problem of interest is

$$\text{minimize } I(P \mid R) \text{ subject to } \int_\Omega \theta \, dP \in C, P \in \mathcal{P}(\Omega) \tag{P}$$

Let $(Z_i)_{i\geq 1}$ be an R-iid sequence on Ω. Problem (P) corresponds to the CLLN for the empirical measures $L_n = \frac{1}{n}\sum_{i=1}^n \delta_{Z_i} \in \mathcal{P}(\Omega)$ with a sequence of conditioning events $\{\frac{1}{n}\sum_{i=1}^n \theta(Z_i) \in C_\delta\}$.

Clearly, $\int_\Omega \log\left(\frac{dP}{dR}\right) dP = \int_\Omega \lambda^*\left(\frac{dP}{dR}\right) dR$ with $\lambda^*(t) = \begin{cases} t\log t - t + 1 & \text{if } t > 0 \\ 1 & \text{if } t = 0 \\ +\infty & \text{if } t < 0 \end{cases}$ which is the convex conjugate of $\lambda(s) = e^s - 1$. This corresponds to $\gamma(s) = e^s - s - 1$, $\gamma^*(t) = \lambda^*(t+1)$, $\gamma_o = \tau$ and $(\gamma_o)^* = \tau^*$ with $\tau(s) \triangleq e^{|s|} - |s| - 1$ and $\tau^*(t) \triangleq (|t|+1)\log(|t|+1) - |t|$.

The corresponding Orlicz spaces are

$$\begin{aligned} L_{\tau^*} &= \{f:\Omega\to\mathbb{R}; \int_\Omega |f|\log|f|\,dR<\infty\} \\ M_\tau &= \{u:\Omega\to\mathbb{R}; \forall\alpha>0, \int_\Omega e^{\alpha|u|}\,dR<\infty\} \\ L_\tau &= \{u:\Omega\to\mathbb{R}; \exists\alpha>0, \int_\Omega e^{\alpha|u|}\,dR<\infty\} \end{aligned}$$

with $M'_\tau = L_{\tau^*}$ and $L'_\tau = L_{\tau^*}\oplus L^s_\tau$. The extended entropy is defined by

$$\bar{I}(\ell\mid R) = I(\ell^a\mid R)+\sup\{\langle\ell^s,u\rangle; u, \int_\Omega e^u\,dR<\infty\},\quad \ell\in\mathcal{E}(\Omega) \tag{5.1}$$

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

$$\mathcal{E}(\Omega)=\{\ell\in L'_\tau; \ell\geq 0, \langle\ell,\mathbf{1}\rangle=1\}.$$

The extended minimization problem is

$$\textit{minimize } \bar{I}(\ell\mid R) \textit{ subject to } \langle\theta,\ell\rangle\in C, \ell\in\mathcal{E}(\Omega). \tag{$\bar{P}$}$$

It is proved in (Léonard and Najim, [23]) that this minimization problem also corresponds to the CLLN for the empirical measures $L_n=\frac{1}{n}\sum_{i=1}^n \delta_{Z_i}\in\mathcal{E}(\Omega)$ with a sequence of conditioning events $\{\frac{1}{n}\sum_{i=1}^n\theta(Z_i)\in C_\delta\}$. But (P) arises when θ defines a good constraint (see (5.4) below) while $(\bar{P})$ arises when θ defines a bad constraint (see (5.6) below).

Note that $\mathcal{E}(\Omega)$ depends on R and that for all $\ell\in\mathcal{E}(\Omega)$, $\ell^a\in\mathcal{P}(\Omega)\cap L_{\tau^*}$: it is a probability measure which is absolutely continuous with respect to R with $\frac{d\ell^a}{dR}$ in $L_{\tau^*}(\Omega,R)$.

The constraint function θ is supposed to be such that for all $y\in\mathcal{Y}$, $\langle y,\theta\rangle$ is measurable and $\langle y,\theta\rangle=0$ R-a.e. if and only if $y=0$.

We introduce the Cramér transform of the image law of R by θ on $\mathcal{X}$:

$$\Xi(x)=\sup_{y\in\mathcal{Y}}\left\{\langle y,x\rangle-\log\int_\Omega e^{\langle y,\theta\rangle}\,dR\right\}\in[0,\infty],\quad x\in\mathcal{X} \tag{5.2}$$

and consider its effective domain: $\mathrm{dom}\,\Xi=\{x\in\mathcal{X};\Xi(x)<\infty\}$.

Proposition 5.3 (Relative entropy subject to good constraints)**.** *Let us assume that θ satisfies the "good constraint" assumption*

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

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

Then, the following dual equality holds:

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

Suppose that in addition $C\cap\mathrm{dom}\,\Xi$ is nonempty. Then, the minimization problem (P) has a unique solution P_* in $\mathcal{P}(\Omega)$, P_* is the entropic projection of R on $\{P\in\mathcal{P}(\Omega), \int_\Omega\theta\,dP\in C\}$ and any minimizing sequence of (P) $\sigma(L_{\tau^*},M_\tau)$-converges to P_*

Suppose that in addition, $C\cap\mathrm{icordom}\,\Xi$ is nonempty, then there exists some linear form y_* on $\mathcal{X}$ such that $\langle y_*,\theta\rangle$ is measurable and

(a) $x_* \triangleq \int_\Omega \theta \, dP_* \in C \cap \mathrm{dom}\, \Xi$
(b) $\langle y_*, x_* \rangle \le \langle y_*, x \rangle, \forall x \in C \cap \mathrm{dom}\, \Xi$
(c) $P_*(d\omega) = \exp\big(\langle y_*, \theta(\omega)\rangle - \log \int_\Omega e^{\langle y_*, \theta\rangle} \, dR\big) \, R(d\omega)$.

In this situation, x_* minimizes Ξ on C, $I(P_* \mid R) = \Xi(x_*)$ and because of (a) and (c), we have

$$x_* = \int_\Omega \theta(\omega) \exp\left(\langle y_*, \theta(\omega)\rangle - \log \int_\Omega e^{\langle y_*, \theta\rangle} \, dR\right) R(d\omega)$$

in the weak sense.

Proposition 5.5 (Relative entropy subject to bad constraints)**.** *Let us assume that θ satisfies the "bad constraint" assumption*

$$\forall y \in \mathcal{Y}, \exists \alpha > 0, \int_\Omega e^{\alpha |\langle y, \theta(\omega)\rangle|} \, R(d\omega) < \infty \tag{5.6}$$

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

Then, the following dual equality holds:

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

Suppose that in addition $C \cap \mathrm{dom}\, \Xi$ is nonempty. Then, the minimization problem $(\bar{P})$ is attained in $\mathcal{E}(\Omega)$: 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 $P_* \in \mathcal{P}(\Omega) \cap L_{\tau^*}$.

Suppose that in addition $C \cap \mathrm{icordom}\, \Xi$ is nonempty. Then , there exists a linear form y_* on $\mathcal{X}$ such that $\langle y_*, \theta \rangle$ is measurable, $\int_\Omega e^{\langle y_*, \theta \rangle} \, dR < \infty$ and

$$P_*(d\omega) = \exp\left(\langle y_*, \theta(\omega)\rangle - \log \int_\Omega e^{\langle y_*, \theta\rangle} \, dR\right) R(d\omega).$$

In Proposition 5.3, x_* is the dominating point in the sense of Ney (see Definition 6.1) of C with respect to Ξ. The representation of x_* has already been obtained for C with a nonempty topological interior in $\mathbb{R}^d$ by Ney in [26] and in a Banach space setting by Einmahl and Kuelbs in [10]. The representation of the I-projection P_* is obtained with a very different proof by Csiszár [6] and ([7], Thm 3). Proposition 5.5 also extends corresponding results of Kuelbs [17] which are obtained in a Banach space setting.
For more details about the minimizers of $(\bar{P})$, one can look at ([21], Theorem 3.4) where a characterization is obtained under the weakest assumption: $C \cap \mathrm{dom}\, \Xi$ is nonempty.

Proof of Propositions 5.3 and 5.5. They are direct consequences of Theorem 3.4, Proposition 4.3, Theorem 4.5 and Lemma 5.7 below. Lemma 5.7 allows to apply our previous results with $\lambda(s) = e^s - 1$ and the extended constraint $\langle (\mathbf{1}, \theta), \ell \rangle \in \{1\} \times C$: the first component of the constraint insures the unit mass: $\langle \mathbf{1}, \ell \rangle = 1$.
The last statement of Proposition 5.5 is proved in ([21], Theorem 3.4). □

Lemma 5.7. *For all $x \in \mathcal{X}$,*

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

Proof. Using the identity: $-\log b = \sup_a\{a+1-be^a\}$, one gets:

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

□

To proceed one step further, we take advantage of a dual equality for the relative entropy which is proved by Csiszár in [7]. At Theorem 5.9 below, the finite dimensional constraint assumption of Theorem 4.7 is removed.
It is assumed that $\mathcal{X}$ is a Banach space. Let $\mathcal{X}'$ denote its topological dual space.

Proposition 5.8 (Csiszár)**.** *Let C be a convex subset of the Banach space $\mathcal{X}$ such that* $\operatorname{int} C\cap \operatorname{dom}\Xi \neq \emptyset$. *Then,*

$$\inf\left\{I(P\mid R); P\in\mathcal{P}(\Omega), \int_\Omega \theta\,dP\in C\right\} = \sup_{y\in\mathcal{X}'}\left\{\inf_{x\in C}\langle y,x\rangle - \log\int_\Omega e^{\langle y,\theta\rangle}\,dR\right\}$$

Proof. This result is a slight modification of (Csiszár, [7], (2.30)). □

Notice that the requirement $\operatorname{int} C\cap\operatorname{dom}\Xi\neq\emptyset$ is more demanding than $C\cap\operatorname{icordom}\Xi\neq\emptyset$ since C is supposed to have a nonempty interior.

Theorem 5.9 (Generalized entropic projection for the relative entropy)**.** *We assume that $\mathcal{X}$ is a Banach space, θ satisfies the "bad constraint" assumption*

$$\forall y\in\mathcal{X}', \exists\alpha>0, \int_\Omega e^{\alpha|\langle y,\theta(\omega)\rangle|}\,R(d\omega)<\infty$$

and C is a closed convex subset of $\mathcal{X}$ such that $\operatorname{int} C\cap\operatorname{dom}\Xi$ *is nonempty.*
Then, $\inf(P)=\inf(\bar{P})$ *and any minimizing sequence of (P) converges in the sense of the $\sigma(L_{\tau^*},M_\tau)$-topology to P_*, the absolutely continuous part of the minimizers.*
Therefore, P_* is the generalized entropic projection of R on $\mathcal{C}=\{P\in\mathcal{P}(\Omega);\int_\Omega\theta\,dP\in C\}$.

Proof. Because of Lemma 5.7, as in the proofs of Propositions 5.3 and 5.5, one can apply the results of the previous sections with $\lambda(s)=e^s-1$. In particular, one can apply Proposition 4.10 since the requirement $\inf(P)=\inf(\bar{P})$ holds true, thanks to Propositions 5.8 and 5.5. □

Remark. Under the assumptions of Theorem 5.3 in the Banach space setting of Theorem 5.9, if $\operatorname{int} C\cap\operatorname{dom}\Xi$ is nonempty, then y_* belongs to $\mathcal{X}'$. It is a consequence of Hahn-Banach theorem since x_* is a supporting point of the convex set C with a non-empty topological interior.

In [7], Csiszár gives an interesting example where the generalized entropic projection P_* can be explicitly computed in a situation where problem (P) is not attained. A detailed analysis of this example in terms of singular component is performed by Léonard and Najim in [23].

6. Dominating points

We are going to investigate some relations between dominating points and entropic projections. In the case where the constraint is good, Theorem 3.7 and Proposition 4.3 state that the minimizer $\bar{x}$ is the dominating point of C and that the generalized entropic projection $Q_* = \bar{Q}$ and $\bar{x}$ are related by the identity:

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

We now look at the situation where the constraint is bad. As remarked in (Léonard and Najim, [23]), an example of Csiszár [7] shows that the above equality may fail. Nevertheless, one still keeps $\bar{x} = \langle \theta, \bar{\ell} \rangle$ where $\bar{\ell}$ is any minimizer of $(\bar{P})$. A necessary and sufficient condition (in terms of the function Λ^*) for $\bar{x}$ to satisfy $\bar{x} = \langle \theta, Q_* \rangle$ is obtained at Theorem 6.6.

Following Ney [25], [26], let us introduce the following definition.

Definition 6.1. *The point $\bar{x}$ in $\mathcal{X}$ is called a* dominating point in the sense of Ney *of the convex and $\sigma(\mathcal{X}, \mathcal{Y})$-closed subset C of $\mathcal{X}$ with respect to the Cramér rate function Ξ (see (5.2)) if*

(a) $\bar{x} \in C \cap \operatorname{dom} \Xi$
(b) *there exists some linear form $\bar{y}$ on $\mathcal{X}$ such that $\langle \bar{y}, \bar{x} \rangle \le \langle \bar{y}, x \rangle$, for all $x \in C \cap \operatorname{dom} \Xi$ and*
(c) $\bar{x} = \displaystyle\int_\Omega \theta(\omega) \frac{\exp(\langle \bar{y}, \theta(\omega) \rangle)}{Z(\bar{y})} \, R(d\omega)$ *where $Z(\bar{y})$ is the unit mass normalizing constant.*

Note that this definition is slightly different from the ones proposed by Ney [26] or Einmahl and Kuelbs [10] since C is neither supposed to be an open set nor to have a non-empty interior and $\bar{x}$ is not assumed to be a boundary point of C. The above integral representation (c) is the integral representation (c) of Proposition 5.3.

Let us call a point $\bar{x} \in \mathcal{X}$ sharing the properties (a), (b) and (3.8) of Theorem 3.7 a *dominating point.*

Definition 6.2. *The point $\bar{x}$ in $\mathcal{X}$ is called a* dominating point *of the convex and $\sigma(\mathcal{X}, \mathcal{Y})$-closed subset C of $\mathcal{X}$ with respect to the rate function Λ^* if*

(a) $\bar{x} \in C \cap \operatorname{dom} \Lambda^*$
(b) *there exists some linear form $\bar{y}$ on $\mathcal{X}$ such that $\langle \bar{y}, \bar{x} \rangle \le \langle \bar{y}, x \rangle$ for all $x \in C \cap \operatorname{dom} \Lambda^*$,*
(c) *$\langle \bar{y}, \theta(\cdot) \rangle$ is measurable and $\bar{x} = \int_\Omega \theta(\omega) \lambda'(\omega, \langle \bar{y}, \theta(\omega) \rangle) \, R(d\omega)$.*

Let us first fix some notations. Recall that the extended entropy $\bar{I}$ is given by (4.4): $\bar{I}(\ell) = I(\ell^a) + I^s(\ell^s)$, $\ell = \ell^a + \ell^s \in L'_{\gamma_o}$. Let us define for all $x \in \mathcal{X}$,

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

Because of the decomposition $L'_{\gamma_o} \simeq L_{\gamma_o^*} \oplus L^s_{\gamma_o}$, one obtains for all $x \in \mathcal{X}$,

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

where $J\Box J^s$ is the inf-convolution of J and J^s. By Theorem 4.5, if $\bar{J}(x) < \infty$, there exists $\ell_x \in L'_{\gamma_o}$ such that $\langle\theta, \ell_x\rangle = x$ and $\bar{J}(x) = \bar{I}(\ell_x)$. Let us define

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

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 $\{\ell \in L'_{\gamma_o}; \langle\theta, \ell\rangle = x\}$ (see Lemma 4.9-(b)). Of course, we have

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

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

Proposition 6.3. *For all $x \in \mathcal{X}$ such that $\bar{J}(x) < \infty$, 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. Let us introduce the *recession function* of Λ^*, defined for all x by

$$\widetilde{\Lambda^*}(x) \triangleq \lim_{t\to+\infty} \Lambda^*(tx)/t \in (-\infty, +\infty].$$

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

Proposition 6.4. *Under the bad constraint assumptions (stated before Theorem 4.5), we have*

$$J^s = \widetilde{\Lambda^*}$$

and the dual equalities:

$$\inf\{\bar{I}(\ell); \ell \in L'_{\gamma_o}, \langle\theta, \ell\rangle = x\} = \Lambda^*(x), \text{ for all } x \in \mathcal{X}$$

and

$$\inf\{I(\ell); \ell \in L_{\gamma_o^*}, \langle\theta, \ell\rangle = x\} = \Lambda^*(x), \text{ for all non-recessive } x \in \mathcal{X}.$$

Proof. Under the bad constraint assumption, by Theorem 4.5 we have: $\bar{J}(x) = \Lambda^*(x)$ and by ([20], Thm 2.3), we get: $J^s(x) = \chi^*(x)$ for all $x \in \mathcal{X}$, where $\chi(y)$ is the convex indicator of dom Λ i.e. $\chi(y) = 0$ if $\Lambda(y) < \infty$ and $+\infty$ otherwise. In other words, χ^* is the support function of dom Λ. Therefore, it is also the recession function of Λ^*. Hence, we have $\Lambda^* = \bar{J} = J\Box J^s = J\Box\widetilde{\Lambda^*}$.
Comparing $\Lambda^* = J\Box\widetilde{\Lambda^*}$ with the general identity $\Lambda^* = \Lambda^*\Box\widetilde{\Lambda^*}$, one obtains that $J(x) = \Lambda^*(x)$, for all non-recessive $x \in \mathcal{X}$. $\Box$

Proposition 6.5. *Under the bad constraint assumptions, for all $x \in \mathcal{X}$ such that $\Lambda^*(x) < \infty$, we have:*

$$\Lambda^*(x) = \Lambda^*(x^a) + \widetilde{\Lambda^*}(x - x^a).$$

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

Proof. By (4.4), we have $\bar{I}(\ell_x) = I(\ell_x^a) + \widetilde{I}(\ell_x - \ell_x^a)$ where $\widetilde{I}$ is the recession function of I. It follows that $\bar{J}(x) = J(x_a) + J^s(x - x^a)$, since $J(x_a) = I(\ell_x^a)$ (Proposition 6.3) and the recession function of $\bar{J}$ is J^s. To see this, note that

- 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 $\bar{J} = \Lambda^*$. The same set of arguments also yields the second statement. □

Theorem 6.6. *Under the bad constraint assumption, let us also assume that $C \cap \operatorname{icordom} \Lambda^*$ is nonempty.*

(a) *Then, a minimizer $\bar{x}$ of Λ^* on the set C is a 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 $(\bar{P})$ are absolutely continuous with respect to R." In such a case the solution of $(\bar{P})$ is unique and it matches the solution of (P).*

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

Proof. Under the assumption that $C \cap \operatorname{icordom} \Lambda^*$ is non-empty, one can apply Theorem B.2 to obtain (a) and (b) of Definition 6.2.
The representation (c) also follows from this theorem. As in ([21], Thm 3.4) where the corresponding function $\bar{\Phi}$ is clarified, one can prove that provided that x stands in $\operatorname{icordom} \Lambda^*$, any minimizer ℓ_x of $\bar{I}$ subject to the constraint $\langle \theta, \ell \rangle = x$, $\ell \in L'_{\gamma_o}$ satisfies

$$\ell_x \in \lambda'(\langle y_x, \theta \rangle) \cdot R + K \tag{6.7}$$

where y_x is a measurable linear form on $\mathcal{X}$ and K is some convex cone of $L^s_{\gamma_o}$. Note that as it is assumed that $x \in \operatorname{icordom} \Lambda^*$, no infinite force field (see [19] for this notion) enters the dual representation of ℓ_x. This means that the absolutely continuous part of ℓ_x is $\ell_x^a = \lambda'(\langle y_x, \theta \rangle) \cdot R$.
Let $\bar{x}$ be a minimizer of Λ^* subject to the constraint $x \in C$. Considering the associated minimizer(s) $\ell_{\bar{x}}$, it follows that if $\bar{x}$ is such that $\bar{x}^s = 0$, then it is a dominating point of C with respect to Λ^*. □

In (Kuelbs, [17], Thm 1), with the setting of Section 5 where $\mathcal{X}$ 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 dominating point in the sense of Ney for *all* convex sets C with a nonempty topological interior is equivalent to some property of the Gâteaux derivative of the log-Laplace transform $y \in \mathcal{X}' \mapsto \log \int_{\mathcal{X}} \exp(\langle y, x \rangle)\, R \circ \theta^{-1}(dx)$ 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 Ξ admits a degenerate recession function.

By (6.7) in the proof of Theorem 6.6, it appears that

$$\ell_{\bar{x}} = \lambda'(\langle \bar{y}, \theta \rangle) \cdot R + \ell^s_{\bar{x}}$$

where $\ell^s_{\bar{x}}$ is a singular linear form. If the gap $\bar{x} - \bar{x}^a = \bar{x}^s = \langle \ell^s_{\bar{x}}, \theta\rangle$ is non-zero, then $\ell^s_{\bar{x}} \neq 0$ is not even a measure in the general case. Its representation in terms of θ and Λ is obtained in [21].

7. The proof of Theorem 2.3

Theorem 2.3 is a restatement of Propositions 7.1, 7.3, 7.5 and 7.7 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 7.1. *Under the assumptions of Theorem 2.3, for all closed subset F of $\mathcal{X}$, 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

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

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,

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

where the last inequality follows from the assumption (d) on the conditioning event. One completes the proof with the following lemma. □

Lemma 7.2. *For any closed set F, $\lim_\delta J(F \cap \mathrm{cl}\ C_\delta) = J(F \cap C) \in [0, \infty]$.*

Proof. As $C \subset \mathrm{cl}\ C_\delta$, for all $\delta > 0$, we have $J(F \cap \mathrm{cl}\ C_\delta) \leq J(F \cap C)$. Since C_δ is nonincreasing, $J(F \cap \mathrm{cl}\ C_\delta)$ is nondecreasing and $\lim_\delta J(F \cap \mathrm{cl}\ C_\delta) = \sup_\delta J(F \cap \mathrm{cl}\ C_\delta) \in [0,\infty]$. If $\sup_\delta J(F \cap \mathrm{cl}\ C_\delta) = \infty$, the inequality $J(F \cap \mathrm{cl}\ C_\delta) \leq J(F \cap C)$ leads to the desired result.
Now, let us suppose that $\sup_\delta J(F \cap \mathrm{cl}\ C_\delta) < \infty$. As $F \cap \mathrm{cl}\ C_\delta$ is closed and J is inf-compact, for any δ there exists $x_\delta \in F \cap \mathrm{cl}\ C_\delta$ such that $J(x_\delta) = J(F \cap \mathrm{cl}\ C_\delta)$ and one can extract a converging subsequence $x_k \to x_*$. Because the C_δ's are nonincreasing, we get $\cap_\delta \mathrm{cl}\ C_\delta = \cap_k \mathrm{cl}\ C_{\delta_k}$ and $\lim_\delta J(F \cap \mathrm{cl}\ C_\delta) = \lim_k J(x_k)$. More, $x_* \in F \cap (\cap_k \mathrm{cl}\ C_{\delta_k}) = F \cap C$ and as J is lsc: $\lim_k J(x_k) \geq J(x_*) \geq J(F \cap C)$. Therefore, $\lim_\delta J(F \cap \mathrm{cl}\ C_\delta) \geq J(F \cap C)$ which completes the proof. □

Let us state the lower bound corresponding to Proposition 7.1.

Proposition 7.3. *If the assumption (d) on the conditioning event is restricted to (d-2):*

$$C \subset \mathrm{int}\ C_\delta, \forall \delta > 0 \tag{7.4}$$

then, for all open subset G of $\mathcal{X}$, 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 $\delta>0$,

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

One concludes with Lemma 7.2. □

Let us recall that $\mathcal{H}=\mathrm{argmin}J_{C_o}$ 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 7.1, we have the following CLLN which is the part of the statement of Theorem 2.3 concerning X_n.

Proposition 7.5. *For all open subset H of $\mathcal{X}$ such that $\mathcal{H}\subset H$, we have*

$$\limsup_{\delta\to 0}\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{P}(X_n\not\in H\mid X_n\in C_\delta)<0.$$

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 $C_\delta\subset\mathcal{X}$. Let us denote $A_\delta\triangleq T^{-1}(C_\delta)=\{\ell\in\mathcal{L}; T\ell\in C_\delta\}$ 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, one has the following transfer result.

Lemma 7.6. *We assume that T is continuous.*

(a) *If C is closed and $J(C)=J(\mathrm{int}\ C)$, then A is closed and $I(A)=I(\mathrm{int}\ A)$.*
(b) *If $C\triangleq\cap_\delta \mathrm{cl}\ C_\delta\subset\mathrm{int}\ C_\delta$ for all $\delta>0$, then $A=\cap_\delta\mathrm{cl}\ A_\delta$ and $A\subset \mathrm{int}\ A_\delta$ for all $\delta>0$.*

Proof. Since T is continuous, for any $A'=T^{-1}C'$, we have: $T^{-1}(\mathrm{int}\ C')\subset\mathrm{int}\ A'\subset A'\subset \mathrm{cl}\ A'\subset T^{-1}(\mathrm{cl}\ C')$.
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, $I(A)=J(C)=J(\mathrm{int}\ C)$ (by hypothesis) $=I(T^{-1}(\mathrm{int}\ C)\geq I(\mathrm{int}\ A)$, since $T^{-1}(\mathrm{int}\ C)\subset\mathrm{int}\ A$. But the converse inequality: $I(A)\leq I(\mathrm{int}\ A)$ is clear.
Let us prove (b). We have: $\cap_\delta\mathrm{cl}\ A_\delta\subset T^{-1}(\cap_\delta\mathrm{cl}\ C_\delta)\triangleq A\triangleq T^{-1}(C)\subset T^{-1}(\cap_\delta\mathrm{int}\ C_\delta)$ (by hypothesis) $\subset\cap_\delta\mathrm{int}\ A_\delta\subset\cap_\delta\mathrm{cl}\ A_\delta$. This proves that all these sets are equal, and in particular that $A=\cap_\delta\mathrm{cl}\ A_\delta$. On the other hand, as for any $\delta>0$, $C\subset\mathrm{int}\ C_\delta$, we have $A=T^{-1}(C)\subset T^{-1}(\mathrm{int}\ C_\delta)\subset\mathrm{int}\ A_\delta$. □

Let us recall that $\mathcal{G}$ is the set of the minimizers of I on A. By the above Lemma 7.6, in the situation of the L_n's, Proposition 7.5 becomes the following

Proposition 7.7. *Under our general assumptions, for all open subset G of $\mathcal{L}$ 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 7.6, the A_δ's share the same properties as the C_δ's. In particular, $I(\operatorname{int} A_\delta) \le I(A)$ also holds for them.

Appendix A. Duality of Orlicz spaces

The function $\rho : \mathbb{R} \to [0, +\infty]$ is called a *Young function* if it is convex, even and satisfies $\rho(0) = 0$, $\lim_{s\to\infty}\rho(s) = +\infty$ and there exists $s_o > 0$ such that $0 \le \rho(s_o) < \infty$. Let Ω be an arbitrary set, $\mathcal{A}$ be a σ-field of subsets of Ω and let R be a nonnegative σ-finite measure on $\mathcal{A}$. In this section, all the numerical functions on Ω are $\mathcal{A}$-measurable and R-almost everywhere equal functions are identified.

The *Orlicz space* associated with ρ is defined by: $L_\rho := \{u : \Omega \to \mathbb{R}\ ; \|u\|_\rho < +\infty\}$ with $\|u\|_\rho = \inf\left\{\beta > 0\ ;\ \int_\Omega \rho\left(\frac{|u(\omega)|}{\beta}\right) R(d\omega) \le 1\right\}$. The function $\|\cdot\|_\rho$ is a norm (the Luxemburg norm) and

$$L_\rho = \{u : \Omega \to \mathbb{R}\ ; \exists\lambda_o > 0, \int_\Omega \rho(\lambda_o u)\, dR < \infty\}.$$

A subspace of interest is

$$M_\rho := \{u : \Omega \to \mathbb{R}\ ; \forall\lambda > 0, \int_\Omega \rho(\lambda u)\, dR < \infty\}.$$

Of course: $M_\rho \subset L_\rho$. The function ρ is said to satisfy the Δ_2-condition if

$$\text{there exist } C > 0, s_o \ge 0 \text{ such that } \forall s \ge s_o, \rho(2s) \le C\rho(s) \tag{A.1}$$

If $s_o = 0$, the Δ_2-condition is said to be global. When R is bounded, in order that $M_\rho = L_\rho$, it is enough that ρ satisfies the Δ_2-condition. When R is unbounded, this equality still holds if the Δ_2-condition is global.
Note that if $\rho(s) = \infty$ for some $s > 0$, M_ρ reduces to the null space. If in addition R is bounded, L_ρ is L_∞. On the other hand, when ρ is a finite function, M_ρ contains all the bounded functions.

Duality in Orlicz spaces is intimately linked with the convex conjugacy. The convex conjugate ρ^* of ρ is also a Young function so that one may consider the Orlicz space L_{ρ^*}. A continuous linear form $\ell \in L'_\rho$ is said to be *singular* if for all $u \in L_\rho$, there exists a nonincreasing sequence of measurable sets (A_n) such that $R(\cap_n A_n) = 0$ and for all $n \ge 1$, $\langle \ell, u\mathbf{1}_{\Omega\setminus A_n}\rangle = 0$. Let us denote L^s_ρ the subspace of L'_ρ of all singular forms.

Theorem A.2 (Representation of L'_ρ).). *Let ρ be any Young function. Any $\ell \in L'_\rho$ is uniquely decomposed as*

$$\ell = \ell^a + \ell^s \tag{A.3}$$

with ℓ^a in $L_{\rho^*} \cdot R$ and ℓ^s in L^s_ρ (the space of all continuous L_{ρ^*}-singular forms on L_ρ). This means that L'_ρ is the direct sum $L'_\rho = (L_{\rho^*} \cdot R) \oplus L^s_\rho$.

If ρ satisfies the Δ_2-condition $L'_\rho = L_{\rho^*}\cdot R$, so that L^s_ρ reduces to the null vector space. In the decomposition (A.3), ℓ^a is called the *absolutely continuous* part of ℓ while ℓ^s is its *singular part.*

Proof. For a proof of this result, see (Giner, [15], Theorems 6.4 and 7.2bis), or for an almost complete result in this direction, see (Kozek, [16], Theorem 2.2). □

We denote $I_\rho(f) = \int_\Omega \rho(f)\,dR \in [0,\infty]$ and $I_{\rho^*}(f) = \int_\Omega \rho^*(f)\,dR \in [0,\infty]$. Let A be a subset of a vector space X in duality with Y. Let $\delta(x \mid A)$ denote the convex indicator function of A. Its support function is $\delta^*(y \mid A) = \sup_{x\in X}\{\langle y,x\rangle - \delta(x\mid A)\} = \sup_{x\in A}\langle y,x\rangle, y\in Y$.

Proposition A.4. *Let I^*_ρ be the convex conjugate of I_ρ for the duality (L_ρ, L'_ρ). For any $\ell\in L'_\rho$, $I^*_\rho(\ell) = I_{\rho^*}(\frac{d\ell^a}{dR}) + \delta^*(\ell^s \mid \mathrm{dom}\, I_\rho)$ where $\ell = \ell^a+\ell^s$ is the decomposition (A.3).*

Proof. This result is (Kozek, [16], Thm 2.6) when ρ is a finite Young function, it is (Rockafellar, [29], Thm 1) when $L_\rho = L_\infty$. For the general case, see (Fougères and Giner, [12], Thm 3.2). □

Proposition A.5. *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 M_ρ.*

Proof. This result is (Fougères and Giner, [12], Cor 4.5). □

Appendix B. A convex minimization problem

The aim of this section is to recall for the convenience of the reader the statements of the main results of ([22], Section 3).

Let us consider $\mathcal{U}_0$ a vector space, $\mathcal{L}_0 = \mathcal{U}_0^\sharp$ its algebraic dual space, Φ a $(-\infty,+\infty]$-valued convex function on $\mathcal{U}_0$ and Φ^* its convex conjugate for the duality $\langle \mathcal{U}_0, \mathcal{L}_0\rangle$ which is defined by

$$\Phi^*(\ell) = \sup_{u\in\mathcal{U}_0}\{\langle \ell,u\rangle - \Phi(u)\},\quad \ell\in\mathcal{L}_0.$$

We shall be concerned with the following convex minimization problem

$$\text{minimize } \Phi^*(\ell) \text{ subject to } T\ell\in C,\ \ell\in\mathcal{L}_0 \tag{P_0}$$

where C is a convex subset of a vector space $\mathcal{X}$ and $T:\mathcal{L}_0\to\mathcal{X}$ is a linear operator.

It is convenient to describe T by means of its adjoint T^T. Let $\mathcal{Y}$ be a vector space in separating duality with $\mathcal{X}$. The operator T is then defined for all $\ell\in\mathcal{L}_0$ and $x\in\mathcal{X}$ by $T\ell = x$ if and only if for all $y\in\mathcal{Y}$, $\langle T^Ty, \ell\rangle_{\mathcal{U}_0,\mathcal{L}_0} = \langle y,x\rangle$. Note that one must assume that

$$T^Ty\in\mathcal{U}_0,\ \forall y\in\mathcal{Y} \tag{B.1}$$

for this definition to be meaningful.
The requirement (B.1) allows us to define $\Gamma(y) = \Phi(T^Ty), y\in\mathcal{Y}$. Its convex conjugate is $\Gamma^*(x) = \sup_{y\in\mathcal{Y}}\{\langle y,x\rangle - \Gamma(y)\}$, $x\in\mathcal{X}$ and the dual problem to (P_0) is

$$\text{maximize } \inf_{x\in C}\langle y,x\rangle - \Gamma(y), y\in\mathcal{Y}. \tag{D_0}$$

A lot of notations. The gauge functional on $\mathcal{U}_0$ of the set $\{u\in\mathcal{U}_0; \max(\Phi(u),\Phi(-u))\le 1\}$ is $|u|_\Phi \stackrel{\Delta}{=} \inf\{\alpha>0; \max(\Phi(u/\alpha),\Phi(-u/\alpha))\le 1\}, u\in\mathcal{U}_0$. Similarly, the gauge functional on $\mathcal{Y}$ of the set $\{y\in\mathcal{Y}; \max(\Gamma(y),\Gamma(-y))\le 1\}$ is $|y|_\Gamma \stackrel{\Delta}{=} \inf\{\alpha>0; \max(\Gamma(y/\alpha),\Gamma(-y/\alpha))\le$

$1\}, y \in \mathcal{Y}$. Under our assumptions (H) below, these are norms. Let $\mathcal{L}_1 \triangleq (\mathcal{U}_0, |\cdot|_\Phi)'$ be the topological dual space of $(\mathcal{U}_0, |\cdot|_\Phi)$ and let $\mathcal{U}_1$ be the $|\cdot|_\Phi$-completion of $\mathcal{U}_0$. Of course, we have $(\mathcal{U}_1, |\cdot|_\Phi)' = \mathcal{L}_1 \subset \mathcal{L}_0$. Similarly, let $\mathcal{X}_1 \triangleq \mathcal{X} \cap (\mathcal{Y}, |\cdot|_\Gamma)'$ be the space of $|\cdot|_\Gamma$-continuous elements of $\mathcal{X}$ and let $\mathcal{Y}_1$ be the $|\cdot|_\Gamma$-completion of $\mathcal{Y}$. We have also $(\mathcal{Y}_1, |\cdot|_\Gamma)' = \mathcal{X}_1 \subset \mathcal{X}$. We denote $C_1 = C \cap \mathcal{X}_1$. Let us denote $\mathcal{Y}_2 = \mathcal{X}_1^\sharp$ the algebraic dual space of $\mathcal{X}_1$ and consider $\bar{\Gamma}(y) \triangleq \sup_{x \in \mathcal{X}_1}\{\langle y, x\rangle - \Gamma^*(x)\}, y \in \mathcal{Y}_2$ the greatest convex $\sigma(\mathcal{Y}_2, \mathcal{X}_1)$-lsc extension of Γ to $\mathcal{Y}_2$. Similarly, let us denote $\mathcal{U}_2 = \mathcal{L}_1^\sharp$ the algebraic dual space of $\mathcal{L}_1$. The greatest convex $\sigma(\mathcal{U}_2, \mathcal{L}_1)$-lsc extension of Φ is $\bar{\Phi}(u) \triangleq \sup_{\ell \in \mathcal{L}_1}\{\langle \ell, u\rangle - \Phi^*(\ell)\}, u \in \mathcal{U}_2$.
The main abstract results of ([22], Section 3) are summarized in Theorem B.2 below. Let us give its underlying hypotheses.

Hypotheses (H)

(H_Φ) 1- $\Phi : \mathcal{U}_0 \to [0, +\infty]$ is convex and $\Phi(0) = 0$
2- $\forall u \in \mathcal{U}_0, \exists \alpha > 0, \Phi(\alpha u) < \infty$
3- $\forall u \in \mathcal{U}_0, u \neq 0, \exists t \in \mathbb{R}, \Phi(tu) > 0$
(H_T) 1- $T^T(\mathcal{Y}) \subset \mathcal{U}_0$
2- $\ker T^T = \{0\}$
(H_C) $C_1 \triangleq C \cap \mathcal{X}_1$ is a convex $\sigma(\mathcal{X}_1, \mathcal{Y}_1)$-closed subset of $\mathcal{X}_1$

Theorem B.2. *Let us assume that the hypotheses* (H) *are fulfilled.*

(a) *The following* little dual equality *holds*

$$\inf\{\Phi^*(\ell); T\ell = x, \ell \in \mathcal{L}_0\} = \Gamma^*(x) \in [0, +\infty], \forall x \in \mathcal{X} \tag{B.3}$$

(b) *We have the following dual equalities*

$$\inf(P_0) = \sup(D_0) = \inf_{x \in C} \Gamma^*(x) \in [0, +\infty].$$

(c) *If* $\inf(P_0) < \infty$, *then* (P_0) *is attained in* $\mathcal{L}_1$. *Moreover, any minimizing sequence for* (P_0) *has* $\sigma(\mathcal{L}_1, \mathcal{U}_1)$*-cluster points and every such cluster point solves* (P_0).

If in addition the following geometrical constraint qualification

$$C \cap \operatorname{icordom} \Gamma^* \neq \emptyset$$

is satisfied[1], *then there exists* $(\bar{\ell}, \bar{y}) \in \mathcal{L}_1 \times \mathcal{Y}_2$ *a solution to* (P_0) *and to the following extended dual problem*

$$\textit{maximize } \inf_{x \in C_1} \langle y, x\rangle - \bar{\Gamma}(y), \quad y \in \mathcal{Y}_2 \tag{D_2}$$

Moreover, $(\bar{\ell}, \bar{y}) \in \mathcal{L}_0 \times \mathcal{Y}_2$ *is a solution to* (P_0) *and* (D_2) *if and only if*

(a) $\bar{x} \triangleq T\bar{\ell} \in C_1$
(b) $\langle \bar{y}, \bar{x}\rangle \leq \langle \bar{y}, x\rangle, \forall x \in C_1$.
(c) $\bar{\ell} \in \partial\bar{\Phi}(T^*\bar{y})$[2]

In this situation, $\bar{x}$ *minimizes* Γ^* *on* C *and we also have* $\bar{x} \in \partial\bar{\Gamma}(\bar{y})$ *and*

$$\Phi^*(\bar{\ell}) + \bar{\Phi}(T^*\bar{y}) = \langle \bar{x}, \bar{y}\rangle = \Gamma^*(\bar{x}) + \bar{\Gamma}(\bar{y}). \tag{B.4}$$

[1]One has $\operatorname{dom} \Gamma^* \subset \mathcal{X}_1$, so that $C_1 \cap \operatorname{dom} \Gamma^* = C \cap \operatorname{dom} \Gamma^*$.
[2]$T^* : \mathcal{Y}_2 \to \mathcal{L}_0^\sharp$ is the natural extension of T^T.

References

[1] R.A. Adams. *Sobolev Spaces.* Academic Press, (1975).

[2] R. Azencott. *Grandes déviations et applications.* in "Ecole d'Eté de Probabilités de Saint-Flour VIII-1978", edited by P.L. Hennequin. Lecture Notes in Mathematics 774, Springer, Berlin, (1980), 1-176.

[3] C. Boucher, R.S. Ellis, B. Turkington. *Derivation of Maximum Entropy Principles in Two-Dimensional Turbulence via Large Deviations.* Journal of Statistical Physics, 98(5/6), 2000, 1235-1277.

[4] N. Bourbaki. *Eléments de mathématique. Intégration 1-4.* 2nd edition, Hermann, (1961).

[5] P. Cattiaux, F. Gamboa. *Large deviations and variational theorems for marginal problems.* Bernoulli 5, 1999.

[6] I. Csiszár. *I-divergence geometry of probability distributions and minimization problems.* Ann. Probab. 3, 1975, 146-158.

[7] I. Csiszár. *Sanov property, generalized I-projection and a conditional limit theorem.* Ann. Probab. 12, 1984, 768-793.

[8] I. Csiszár. *Generalized projections for non-negative functions.* Acta Math. Hungar. 68, 1995, 161-185.

[9] A. Dembo, O. Zeitouni. *Large Deviations Techniques and Applications*, Applications of Mathematics 38, Springer (1998).

[10] U. Einmahl, J. Kuelbs. *Dominating points and large deviations for random vectors.* Probab. Theory Relat. Fields 105, 1996, 529-543.

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

[12] A. Fougères, E. Giner. *Applications de la décomposition d'un espace d'Orlicz ...*, Séminaire d'Analyse Convexe. Montpellier 76, Exposé 18, (1976), in French.

[13] M.I. Freidlin, A.D. Wentzell. *Random Perturbation of Dynamical Systems.*, 2nd ed., Springer, New York, 1998.

[14] F. Gamboa, E. Gassiat. *Bayesian methods and maximum entropy for ill-posed inverse problems.* Ann. Stat. 25, 1997, 328-350.

[15] E. Giner. *Décomposabilité et dualité des espaces d'Orlicz.* Séminaire d'analyse convexe, Montpellier 76, exposé 16, 1976.

[16] A. Kozek. *Convex integral functionals on Orlicz spaces.* Ann. Soc. Math. Polonae, Series I, 1979, 109-135.

[17] J. Kuelbs. *Large deviation probabilities and dominating points for open convex sets: nonlogarithmic behavior.* Ann. Probab. 28(3), 2000, 1259-1279.

[18] C. Léonard. *Convex conjugates of integral functionals.* Acta Math. Hungar. 93(4), 2001, 253-280.

[19] C. Léonard. *Minimizers of energy functionals.* Acta Math. Hungar. 93(4), 2001, 281-325.

[20] C. Léonard. *Minimization of energy functionals applied to some inverse problems.* Appl. Math. Optim. 44, 2001, 273-297.

[21] C. Léonard. *Minimizers of energy functionals under not very integrable constraints.* J. Convex Anal. 10(1), 2003, 63-88.

[22] C. Léonard. *A convex optimization problem arising from probabilistic questions. Preprint CMAP-Ecole Polytechnique, 506*, 2003, www.cmap.polytechnique.fr/preprint/

[23] C. Léonard, J. Najim. *An extension of Sanov's theorem: application to the Gibbs conditioning principle.* Bernoulli 8(6), 2002, 721-743.

[24] J. Najim. *A Cramér Type Theorem for Weighted Random Variables.* Electronic Journal of Probability, 7(4), 2002, 1-32.

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

[26] P. Ney. *Convexity and large deviations.* Ann. Probab. 12, 1984, 903-906.

[27] M. Rao, Z.D. Ren. *Theory of Orlicz Spaces.* Pure and Applied Mathematics, 146, Marcel Dekker, (1991).

[28] R.T. Rockafellar. *Integrals which are convex functionals.* Pacific J. Math. 24(3), 1968, 525-539.

[29] R.T. Rockafellar. *Integrals which are convex functionals, II.* Pacific J. Math., 39(2), (1971), 439–469.
[30] R.T. Rockafellar. *Conjugate Duality and Optimization*, Regional Conferences Series in Applied Mathematics, SIAM, Philadelphia (1974).
[31] D.W. Stroock, O. Zeitouni. Microcanonical distributions, Gibbs states and the equivalence of ensembles. In: *Festchrift in Honour of F. Spitzer.* Eds. R. Durrett and H. Kesten, Birkhaüser (1991), 399-424.
[32] M. Teboulle, I. Vajda. Convergence of Best ϕ-Entropy Estimates. *IEEE Trans. Inform. Theory*, 39, (January 1993), 297–301.

Modal-X, Université Paris 10 &

CMAP, École Polytechnique. 91128 Palaiseau Cedex, France
E-mail address: christian.leonard@polytechnique.fr