



HAL
open science

Entropic Projections and Dominating Points

Christian Léonard

► **To cite this version:**

Christian Léonard. Entropic Projections and Dominating Points. ESAIM: Probability and Statistics, 2011, 10.1051/ps/2009003 . hal-00184785v1

HAL Id: hal-00184785

<https://hal.science/hal-00184785v1>

Submitted on 1 Nov 2007 (v1), last revised 7 Sep 2010 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ENTROPIC PROJECTIONS AND DOMINATING POINTS

CHRISTIAN LÉONARD

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

CONTENTS

1. Introduction	1
2. Minimizing entropy under convex constraints	4
3. Some examples	11
4. Solving (P_C) with bad constraints	13
5. Entropic projections	20
6. Dominating points	24
References	27

1. INTRODUCTION

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

Conditional laws of large numbers. Suppose that the empirical measures

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

of the \mathcal{Z} -valued random variables Z_1, Z_2, \dots (δ_z is the Dirac measure at z) obey a Large Deviation Principle (LDP) in the set $P_{\mathcal{Z}}$ of all probability measures on \mathcal{Z} with the rate

Date: November 07.

2000 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, entropic projections, dominating points.

function I . This approximately means that $\mathbb{P}(L_n \in \mathcal{A}) \underset{n \rightarrow \infty}{\asymp} \exp[-n \inf_{P \in \mathcal{A}} I(P)]$ for $\mathcal{A} \subset P_{\mathcal{Z}}$. With regular enough subsets \mathcal{A} and \mathcal{C} of $P_{\mathcal{Z}}$, one can expect that for “all” \mathcal{A}

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

where P_* is a minimizer of I on \mathcal{C} . To see this, remark that (formally) $\mathbb{P}(L_n \in \mathcal{A} \mid L_n \in \mathcal{C}) \underset{n \rightarrow \infty}{\asymp} \exp[-n(\inf_{P \in \mathcal{A} \cap \mathcal{C}} I(P) - \inf_{P \in \mathcal{C}} I(P))]$. If I is strictly convex and \mathcal{C} is convex, P_* is unique and this roughly means that conditionally on $L_n \in \mathcal{C}$, as n tends to infinity L_n tends to the solution P_* of the minimization problem

$$\text{minimize } I(P) \text{ subject to } P \in \mathcal{C}, \quad P \in P_{\mathcal{Z}} \quad (1.2)$$

Such conditional Laws of Large Numbers (LLN) appear in information theory and in statistical physics where they are often called Gibbs conditioning principles (see [7, Section 7.3] and the references therein). If the variables Z_i are independent and identically distributed with law R , the LDP for the empirical measures is given by Sanov’s theorem and the rate function I is the relative entropy

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

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

$$L_n = \frac{1}{n} \sum_{i=1}^n W_i \delta_{z_i} \quad (1.3)$$

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

$$I(Q) = \begin{cases} \int_{\mathcal{Z}} \gamma_z^* \left(\frac{dQ}{dR}(z) \right) R(dz), & \text{if } Q \prec R \\ +\infty, & \text{otherwise} \end{cases}, \quad Q \in M_{\mathcal{Z}} \quad (1.4)$$

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

$$\text{minimize } I(Q) \text{ subject to } Q \in \mathcal{C}, \quad Q \in M_{\mathcal{Z}} \quad (1.5)$$

The large deviations of these random measures and their conditional LLNs enter the framework of Maximum Entropy in the Mean (MEM) which has been studied among others by Dacunha-Castelle, Csiszár, Gamboa, Gassiat, Najim see [5, 6, 10, 21] and also [7, Theorem 7.2.3]. This problem also arises in the context of statistical physics. It has been studied among others by Boucher, Ellis, Gough, Puli and Turkington, see [1, 9].

The relative entropy corresponds to $\gamma^*(t) = t \log t - t + 1$ in (1.4): the problem (1.2) is a special case of (1.5) with the additional constraint that $Q(\mathcal{Z}) = 1$.

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

Entropic projections. The minimizers of (1.5) are called *entropic projections*. It may happen that even if the minimizer is not attained, any minimizing sequence converges to some measure Q_* which does not belong to \mathcal{C} . This intriguing phenomenon was discovered by Csiszár [3]. Such a Q_* is called a *generalized* entropic projection.

In the special case where I is the relative entropy, Csiszár has obtained existence results in [2] together with dual equalities. His proofs are based on geometric properties of the relative entropy; no convex analysis is needed. Based on the same geometric ideas, he obtained later in [3] a powerful Gibbs conditioning principle for noninteracting particles. For general entropies as in (1.4), he studies the problem of existence of entropic and generalized entropic projections in [4].

The minimization problem (1.5) is interesting in its own right, even when conditional LLNs are not at stake. The literature on this subject is huge. Some bibliographical entries are given in [15].

Dominating points. Let the constraint set \mathcal{C} be described by

$$\mathcal{C} = \{Q \in M_{\mathcal{Z}}; TQ \in C\} \quad (1.6)$$

where C is a subset of a vector space \mathcal{X} and $T : M_{\mathcal{Z}} \rightarrow \mathcal{X}$ is a linear operator. As a typical example, one can think of $TQ = \int_{\mathcal{Z}} \theta(z) Q(dz)$ where $\theta : \mathcal{Z} \rightarrow \mathcal{X}$ is some function and the integral should be taken formally for the moment. With L_n given at (1.1) or (1.3), if T is regular enough, we obtain by the contraction principle that $X_n := TL_n = \frac{1}{n} \sum_{i=1}^n \theta(Z_i) \in \mathcal{X}$ or $X_n := TL_n = \frac{1}{n} \sum_{i=1}^n W_i \theta(z_i)$ obeys the LDP in \mathcal{X} with rate function $J(x) = \inf\{I(Q); Q \in M_{\mathcal{Z}}, TQ = x\}$, $x \in \mathcal{X}$. Once again, the conditional LLN for X_n is of the form: For “all” $A \subset \mathcal{X}$,

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

where x_* is a solution to the minimization problem

$$\text{minimize } J(x) \text{ subject to } x \in C, \quad x \in \mathcal{X} \quad (1.7)$$

The minimizers of (1.7) are called *dominating points*. This notion was introduced by Ney [22, 23] in the special case where $(Z_i)_{i \geq 1}$ is an iid sequence in $\mathcal{Z} = \mathbb{R}^d$ and θ is the identity, i.e. $X_n = \frac{1}{n} \sum_{i=1}^n Z_i$. Later, Einmahl and Kuelbs [8, 13] have extended this study to a Banach space \mathcal{Z} . In this iid case, J is the Cramér transform of the law of Z_1 .

Presentation of the results. One treats the problems of existence of entropic projections and dominating points in a unified way, taking advantage of the mapping $TQ = x$. Although it is simple, this connection does not seem to be used in previous literature. Hence, one mainly concentrates efforts on the entropic projections and then transports the results to the dominating points.

It will be proved at Proposition 5.5 that the entropic projection exists on \mathcal{C} if the supporting hyperplanes of \mathcal{C} are directed by sufficiently integrable functions. In some cases of not enough integrable supporting hyperplanes, the representation of the generalized projection is still available and given at Theorem 5.6. It will appear that the generalized projection is the “measure” part of the minimizer of an *extended* minimization problem.

For instance, with the relative entropy $I(\cdot|R)$, the projection exists in \mathcal{C} if its supporting hyperplanes are directed by functions u such that $\int_{\mathcal{Z}} e^{\alpha|u|} dR < \infty$ for all $\alpha > 0$, see

Proposition 5.10. If these u only satisfy $\int_{\mathcal{Z}} e^{\alpha|u|} dR < \infty$ for *some* $\alpha > 0$, the projection may not exist in \mathcal{C} , but the generalized projection is computable: its Radon-Nykodym derivative with respect to R is characterized at Proposition 5.13.

One finds again some already known results of Csiszár [3, 4], U. Einmahl and Kuelbs [8, 13] with different proofs and a new point of view. The representations of the generalized projections are new results. The conditions on C to obtain dominating points are improved and an interesting phenomenon noticed in [13] is clarified at Remark 6.9 by connecting it with the generalized entropic projection.

The main results are Theorems 4.1, 4.7, 5.6 and 6.8.

Outline of the paper. At Section 2, one gives a precise formulation of the entropy minimization problem (1.5) and a natural extension of it, see (\overline{P}_C) at Section 2.3. Then one recalls at Theorems 2.14 and 2.17 results from [15] about the existence and uniqueness of the solutions of (1.5) and (\overline{P}_C) , related dual equalities and the characterizations of their solutions in terms of integral representations.

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

One shows at Theorem 4.7 in Section 4 that under “bad” constraints (this notion is specified at Section 2.4 in terms of the set C and the operator T appearing in (1.6)), although the problem (1.5) may not be attained, its minimizing sequences may converge in some sense to some measure Q_* : the *generalized* entropic projection.

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

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

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

Let $f : X \rightarrow [-\infty, +\infty]$ be an extended numerical function. Its convex conjugate with respect to $\langle X, Y \rangle$ is $f^*(y) = \sup_{x \in X} \{\langle x, y \rangle - f(x)\} \in [-\infty, +\infty]$, $y \in Y$. Its subdifferential at x with respect to $\langle X, Y \rangle$ is $\partial_Y f(x) = \{y \in Y; f(x + \xi) \geq f(x) + \langle y, \xi \rangle, \forall \xi \in X\}$. If no confusion occurs, one writes $\partial f(x)$.

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

The indicator of a subset A of X is defined by

$$\iota_A(x) = \begin{cases} 0, & \text{if } x \in A \\ +\infty, & \text{otherwise} \end{cases}, \quad x \in X.$$

One writes

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

and $I = I_{\gamma^*}$ for short, instead of (1.4).

2. MINIMIZING ENTROPY UNDER CONVEX CONSTRAINTS

In this section, the main results of [15] are recalled.

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

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

Definitions 2.1 (The Orlicz spaces L_ρ and E_ρ). *The Orlicz space associated with ρ is defined by $L_\rho = \{u : \mathcal{Z} \rightarrow \mathbb{R}; \|u\|_\rho < +\infty\}$ where the Luxemburg norm $\|\cdot\|_\rho$ is defined by $\|u\|_\rho = \inf \{\beta > 0 ; \int_{\mathcal{Z}} \rho(z, u(z)/\beta) R(dz) \leq 1\}$ and R -a.e. equal functions are identified. Hence,*

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

A subspace of interest is

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

Of course $E_\rho \subset L_\rho$. Note that if ρ doesn't depend on z and $\rho(s_o) = \infty$ for some $s_o > 0$, E_ρ reduces to the null space and if in addition R is bounded, L_ρ is L_∞ . On the other hand, if ρ is a finite function which doesn't depend on z and R is bounded, E_ρ contains all the bounded functions.

Duality in Orlicz spaces is intimately linked with the convex conjugacy. The convex conjugate ρ^* of ρ is defined by $\rho^*(z, t) = \sup_{s \in \mathbb{R}} \{st - \rho(z, s)\}$. It 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 decreasing sequence of measurable sets (A_n) such that $R(\cap_n A_n) = 0$ and for all $n \geq 1$, $\langle \ell, u \mathbf{1}_{\mathcal{Z} \setminus A_n} \rangle = 0$. Let us denote L_ρ^s the subspace of L'_ρ of all singular forms.

Theorem 2.2 (Representation of E'_ρ and L'_ρ).

(a) *Suppose that ρ is a finite Young function. Then,*

(i) E'_ρ is isomorphic to L_{ρ^*} ;

(ii) $\ell \in L'_\rho$ is singular if and only if $\langle \ell, u \rangle = 0$, for all u in E_ρ .

(b) *Let ρ be any Young function. Any $\ell \in L'_\rho$ is uniquely decomposed as*

$$\ell = \ell^a + \ell^s \tag{2.3}$$

with $\ell^a \in L_{\rho^} R$ and $\ell^s \in L_\rho^s$. This means that L'_ρ is the direct sum $L'_\rho = L_{\rho^*} R \oplus L_\rho^s$.*

We denote the space of R -absolutely continuous signed measures having a density in the Orlicz space L_{ρ^*} by $L_{\rho^*} R$.

Proof. For a proof of (a), see [11, Thm 4.8] and [12, Proposition 2.1].

About (b): When $L_\rho = L_\infty$ this result is the usual representation of L'_∞ . When ρ is a finite function, this result is ([12], Theorem 2.2). The general result is proved in [24], with ρ not depending on z but the extension to a z -dependent ρ is obvious. \square

In the decomposition (2.3), ℓ^a is called the *absolutely continuous* part of ℓ while ℓ^s is its *singular part*. ℓ^a is a measure which is absolutely continuous with respect to R and ℓ^s is not a measure: it is additive but not σ -additive.

Definition 2.4 (Δ_2 -condition). *The function ρ is said to satisfy the Δ_2 -condition if there exist $C > 0$, $s_o \geq 0$ such that for all $s \geq s_o$, $\rho(2s) \leq C\rho(s)$.*

When R is bounded, in order that $E_\rho = L_\rho$, it is enough that ρ satisfies the Δ_2 -condition. If R is unbounded, for $E_\rho = L_\rho$, it is enough that the Δ_2 -condition holds globally, i.e. with $s_o = 0$. Consequently, if ρ satisfies the Δ_2 -condition we have $L'_\rho = L_{\rho^*}R$ so that L_ρ^s reduces to the null vector space.

2.2. The entropy minimization problem (P_C).

Entropy. Let R be a positive measure on a space \mathcal{Z} and take a $[0, \infty]$ -valued measurable function γ^* on $\mathcal{Z} \times \mathbb{R}$ such that $\gamma^*(z, \cdot) := \gamma_z^*$ is convex and lower semicontinuous for all $z \in \mathcal{Z}$. Denote $M_{\mathcal{Z}}$ the space of all signed measures Q on \mathcal{Z} . The entropy functional to be considered is defined by

$$I(Q) = \begin{cases} \int_{\mathcal{Z}} \gamma_z^* \left(\frac{dQ}{dR}(z) \right) R(dz) & \text{if } Q \prec R \\ +\infty & \text{otherwise} \end{cases}, \quad Q \in M_{\mathcal{Z}}. \quad (2.5)$$

where $Q \prec R$ means that Q is absolutely continuous with respect to R . Assume that for each z there exists a unique $m(z)$ which minimizes γ_z^* with

$$\gamma_z^*(m(z)) = 0, \quad \forall z \in \mathcal{Z}.$$

Then, I is $[0, \infty]$ -valued, its unique minimizer is mR and $I(mR) = 0$.

Relevant Orlicz spaces. Since γ_z^* is closed convex for each z , it is the convex conjugate of some closed convex function γ_z . Defining

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

one sees that for R -a.e. z , λ_z is a nonnegative convex function and it vanishes at 0. Hence,

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

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

The effective domain of I is included in $mR + L_{\lambda_\diamond^*}R$. It will be assumed from now on that $m \in L_{\lambda_\diamond^*}$ so that $\text{dom } I = L_{\lambda_\diamond^*}R$.

Constraint. In order to define the constraint, take \mathcal{X}_o a vector space and a function $\theta : \mathcal{Z} \rightarrow \mathcal{X}_o$. One wants to give some meaning to the formal constraint $\int_{\mathcal{Z}} \theta d\ell = x$ with $\ell \in L_{\lambda_\diamond^*}R$ and $x \in \mathcal{X}_o$. Suppose that \mathcal{X}_o is the algebraic dual space of some vector space \mathcal{Y}_o and define for all $y \in \mathcal{Y}_o$ and $z \in \mathcal{Z}$, $\langle y, \theta(z) \rangle_{\mathcal{Y}_o, \mathcal{X}_o}$. Assuming that

$$\langle \mathcal{Y}_o, \theta(\cdot) \rangle \subset L_{\lambda_\diamond}, \quad (2.6)$$

Hölder's inequality in Orlicz spaces allows to define the constraint operator $\ell \in L_{\lambda_\diamond} \mapsto \langle \theta, \ell \rangle \in \mathcal{X}_o$ by: $\langle y, \langle \theta, \ell \rangle \rangle_{\mathcal{Y}_o, \mathcal{X}_o} = \langle \langle y, \theta \rangle, \ell \rangle_{L_{\lambda_\diamond}, L_{\lambda_\diamond^*}}, \forall y \in \mathcal{Y}_o$. If $\ell \in L_{\lambda_\diamond^*}R$, one sometimes write $\langle \theta, \ell \rangle = \int_{\mathcal{Z}} \theta d\ell$. One sometimes denote

$$T\ell = \langle \theta, \ell \rangle, \quad \ell \in L_{\lambda_\diamond^*}.$$

Minimization problem. We are now ready to state the minimization problem (1.5) precisely:

$$\text{minimize } I(Q) \text{ subject to } \int_{\mathcal{Z}} \theta dQ \in C, \quad Q \in L_{\lambda_\diamond^*}R \quad (\text{P}_C)$$

where C is a convex subset of \mathcal{X}_o .

2.3. The extended entropy minimization problem(\bar{P}_C). The extended entropy is defined by

$$\bar{I}(\ell) = I(\ell^a) + I^s(\ell^s), \quad \ell \in L'_{\lambda_\diamond} \quad (2.7)$$

where, using the notation of Theorem 2.2,

$$I^s(\ell^s) = \iota_{\text{dom } I_\gamma}^*(\ell^s) = \sup \{ \langle \ell^s, u \rangle; u \in L_{\lambda_\diamond}, I_\gamma(u) < \infty \} \in [0, \infty].$$

The associated extended minimization problem is

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

\bar{I} is the greatest convex $\sigma(L'_{\lambda_\diamond}, L_{\lambda_\diamond})$ -lower semicontinuous extension of I to $L'_{\lambda_\diamond} \supset L_{\lambda_\diamond^*}R$, see [15] for references about this result.

2.4. Good and bad constraints. If the Young function λ_\diamond doesn't satisfy the Δ_2 -condition (see Definition 2.4), for instance if it has an exponential growth at infinity as in (3.1) below, the *small* Orlicz space E_{λ_\diamond} may be a proper subset of L_{λ_\diamond} . Consequently, for some functions θ , the integrability property

$$\langle \mathcal{Y}_o, \theta(\cdot) \rangle \subset E_{\lambda_\diamond} \quad (2.8)$$

or equivalently

$$\forall y \in \mathcal{Y}_o, \int_{\mathcal{Z}} \lambda(z, \langle y, \theta(z) \rangle) R(dz) < \infty \quad (A_\theta^\forall)$$

may not be satisfied while the weaker property (2.6): $\langle \mathcal{Y}_o, \theta(\cdot) \rangle \subset L_{\lambda_\diamond}$, or equivalently

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

holds. In this situation, analytical complications occur (see Section 4). This is the reason why constraint satisfying (A_θ^\forall) are called *good constraints*, while constraints satisfying (A_θ^\exists) but not (A_θ^\forall) are called *bad constraints*.

One denotes \mathcal{Y}_L the subset of \mathcal{X}_o^* which is isomorphic to the strong closure of the subspace $\{ \langle y, \theta \rangle; y \in \mathcal{Y}_o \}$ in L_{λ_\diamond} . Under the assumption (A_θ^\forall) , for all $y \in \mathcal{Y}_o$, $\langle y, \theta \rangle \in E_{\lambda_\diamond}$ and $\mathcal{Y}_L = \mathcal{Y}_E$: the strong closure of $\{ \langle y, \theta \rangle; y \in \mathcal{Y}_o \}$ in E_{λ_\diamond} .

Under the assumption (A_θ^\forall) , the convex set C is said to be a *good constraint set* if

$$T^{-1}C \cap L_{\lambda_\diamond^*}R = \bigcap_{y \in Y} \left\{ fR \in L_{\lambda_\diamond^*}R; \int_{\mathcal{Z}} \langle y, \theta \rangle f dR \geq a_y \right\} \text{ for some } Y \subset \mathcal{Y}_E \quad (2.9)$$

and some function $y \in Y \mapsto a_y \in \mathbb{R}$. Under the assumption (A_θ^\exists) , it is said to be a *bad constraint set* if

$$T^{-1}C = \bigcap_{y \in Y} \{ \ell \in L'_{\lambda_\diamond}; \langle \langle y, \theta \rangle, \ell \rangle \geq a_y \} \text{ for some } Y \subset \mathcal{Y}_L \quad (2.10)$$

These special shapes (2.9) and (2.10) imply some closure property of C . For comparison, if (A_θ^\exists) holds and C is only supposed to be convex, $T^{-1}C = \bigcap_{(y,a) \in A} \{ \ell \in L'_{\lambda_\diamond}; \langle \langle y, \theta \rangle, \ell \rangle > a \}$ with $A \subset \mathcal{Y}_L \times \mathbb{R}$.

2.5. Several function spaces and cones. To state the extended dual problems (\tilde{D}_C) and (\overline{D}_C) below, notation is needed. If λ is not an even function, one has to consider

$$\begin{cases} \lambda_+(z, s) = \lambda(z, |s|) \\ \lambda_-(z, s) = \lambda(z, -|s|) \end{cases} \quad (2.11)$$

which are Young functions and the corresponding Orlicz spaces.

Let E be a Riesz vector space for the order relation \leq . Any $e \in E$ admits a positive part: $e_+ := e \vee 0$ and a negative part: $e_- := (-e) \vee 0$. One writes $|e| = e_+ + e_-$. There is a natural order on the algebraic dual space E^* of a Riesz vector space E which is defined by: $e^* \leq f^* \Leftrightarrow \langle e^*, e \rangle \leq \langle f^*, e \rangle$ for any $e \in E$ with $e \geq 0$. A linear form $e^* \in E^*$ is said to be *relatively bounded* if for any $f \in E$, $f \geq 0$, we have $\sup_{e: |e| \leq f} |\langle e^*, e \rangle| < +\infty$. Although E^* may not be a Riesz space in general, the vector space E^b of all the relatively bounded linear forms on E is always a Riesz space. In particular, the elements of E^b admit a decomposition in positive and negative parts $e^* = e_+^* - e_-^*$.

Definitions 2.12. For any relatively bounded linear form ζ on L'_{λ_\circ} i.e. $\zeta \in L_{\lambda_\circ}^b$, one writes:

- $\zeta \in K''_\lambda$ to specify that $\zeta_{\pm|L'_{\lambda_\pm} \cap L'_{\lambda_\circ}} \in L''_{\lambda_\pm}$
- $\zeta \in K'_{\lambda^*}$ to specify that $\zeta_{\pm|L_{\lambda_\pm^*} R \cap L'_{\lambda_\circ}} \in L'_{\lambda_\pm^*}$
- $\zeta \in K_\lambda$ to specify that $\zeta_{\pm|L_{\lambda_\pm} R \cap L'_{\lambda_\circ}} \in L_{\lambda_\pm}$
- $\zeta \in K_{\lambda^*}^s$ to specify that $\zeta_{\pm|L_{\lambda_\pm^*} R \cap L'_{\lambda_\circ}} \in L_{\lambda_\pm^*}^s$
- $\zeta \in K_{\lambda'}^{s'}$ to specify that $\zeta_{\pm|L_{\lambda_\pm}^s \cap L'_{\lambda_\circ}} \in L_{\lambda_\pm}^{s'}$

where λ_\pm are defined at (2.11) and $\zeta_{\pm|L_\pm \cap L'_{\lambda_\circ}} \in L'_\pm$ means that the restriction of ζ_\pm to $L_\pm \cap L'_{\lambda_\circ}$ is continuous with respect to relative topology generated by the strong topology of L_\pm on $L_\pm \cap L'_{\lambda_\circ}$.

- (1) The sets $K''_\lambda, K'_{\lambda^*}, K_\lambda, K_{\lambda^*}^s$ and $K_{\lambda'}^{s'}$ are defined to be the corresponding subsets of $L_{\lambda_\circ}^b$. They are not vector spaces in general but convex cones with vertex 0.
- (2) The $\sigma(K''_\lambda, K'_\lambda)$ -closure \overline{A} of a set A is defined as follows: $\zeta \in L_{\lambda_\circ}^b$ is in \overline{A} if $\zeta_{\pm|L'_{\lambda_\pm} \cap L'_{\lambda_\circ}}$ is in the $\sigma(L''_{\lambda_\pm} \cap L_{\lambda_\circ}^b, L'_{\lambda_\pm} \cap L'_{\lambda_\circ})$ -closure of $A_\pm = \{\zeta_\pm; \zeta \in A\}$. Clearly, $\overline{A}_\pm = \{\zeta_\pm; \zeta \in \overline{A}\}$. One defines similarly the $\sigma(K'_{\lambda^*}, K_{\lambda^*})$, $\sigma(K_\lambda, K'_\lambda)$, $\sigma(K_{\lambda^*}^s, K_{\lambda^*})$ and $\sigma(K_{\lambda'}^{s'}, K_{\lambda'}^s)$ -closures.
- (3) Let A be a subset of L_{λ_\circ} . Its strong closure $s\text{-cl } A$ in K_λ is the set of all measurable functions u such that u_\pm is in the $\|\cdot\|_{\lambda_\pm}$ -closure of $A_\pm = \{v_\pm; v \in A\}$.

Let ρ be a Young function. By Theorem 2.2, we have $L''_\rho = [L_\rho \cdot R \oplus L_\rho^s] \oplus L_{\rho'}^{s'}$. For any $\zeta \in L''_\rho = (L_{\rho^*} R \oplus L_\rho^s)'$, let us denote the restrictions $\zeta_1 = \zeta|_{L_{\rho^*} R}$ and $\zeta_2 = \zeta|_{L_{\rho'}^s}$. Since, $(L_{\rho^*} R)' \simeq L_\rho \oplus L_{\rho^*}^s$, one sees that any $\zeta \in L''_\rho$ is uniquely decomposed into

$$\zeta = \zeta_1^a + \zeta_1^s + \zeta_2 \quad (2.13)$$

with $\zeta_1 = \zeta_1^a + \zeta_1^s \in L_{\rho^*}'$, $\zeta_1^a \in L_\rho$, $\zeta_1^s \in L_{\rho^*}^s$ and $\zeta_2 \in L_{\rho'}^{s'}$. With our definitions, $K''_\lambda = [K_\lambda \oplus K_{\lambda^*}^s] \oplus K_{\lambda'}^{s'}$ and the decomposition (2.13) holds for any $\zeta \in K''_\lambda$ with

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

2.6. The assumptions. Let us first collect the assumptions on R, γ^* and θ .

Assumptions (A).

(A_R) It is assumed that the reference measure R is a σ -finite nonnegative measure on a space \mathcal{Z} endowed with some R -complete σ -field.

(A _{γ^*}) *Assumptions on γ^* .*

(1) $\gamma^*(\cdot, t)$ is z -measurable for all t and for R -almost every $z \in \mathcal{Z}$, $\gamma^*(z, \cdot)$ is a lower semicontinuous strictly convex $[0, +\infty]$ -valued function on \mathbb{R} which attains a unique minimum. Let $m(z)$ denote this minimizer.

(2) It is also assumed that for R -almost every $z \in \mathcal{Z}$, the minimum value is $\gamma^*(z, m(z)) = 0$ and there exist $a(z), b(z) > 0$ such that $0 < \gamma^*(z, m(z) + a(z)) < \infty$ and $0 < \gamma^*(z, m(z) - b(z)) < \infty$.

(3) $\int_{\mathcal{Z}} \lambda^*(z, \alpha m(z)) R(dz) + \int_{\mathcal{Z}} \lambda^*(z, -\alpha m(z)) R(dz) < \infty$, for some $\alpha > 0$.

(4) For R -almost every $z \in \mathcal{Z}$, $\lim_{t \rightarrow \pm\infty} \gamma_z^*(t)/t = +\infty$.

(A _{θ}) *Assumptions on θ .*

(1) for any $y \in \mathcal{Y}_o$, the function $z \in \mathcal{Z} \mapsto \langle y, \theta(z) \rangle \in \mathbb{R}$ is measurable;

(2) for any $y \in \mathcal{Y}_o$, $\langle y, \theta(\cdot) \rangle = 0, R$ -a.e. implies that $y = 0$;

(\exists) $\forall y \in \mathcal{Y}_o, \exists \alpha > 0, \int_{\mathcal{Z}} \lambda(z, \alpha \langle y, \theta(z) \rangle) R(dz) < \infty$.

2.7. The solution of (P_C). The dual problem associated with (P_C) is

$$\text{maximize } \inf_{x \in C \cap \mathcal{X}_L} \langle y, x \rangle - I_\gamma(\langle y, \theta \rangle), \quad y \in \mathcal{Y}_L \quad (\text{D}_C)$$

and the extended dual problem associated with (P_C) is

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

where $\tilde{\mathcal{Y}}$ is the convex cone of all linear forms ω on \mathcal{X}_o such that $\langle \omega, \theta(\cdot) \rangle_{\mathcal{X}_o^*, \mathcal{X}_o}$ is in the $\sigma(K_\lambda, K_{\lambda^*})$ -closure of $\{\langle y, \theta \rangle; y \in \mathcal{Y}_o\}$.

Let us define

$$\Gamma^*(x) = \sup_{y \in \mathcal{Y}_o} \left\{ \langle y, x \rangle - \int_{\mathcal{Z}} \gamma(\langle y, \theta \rangle) dR \right\}, \quad x \in \mathcal{X}_o$$

the convex conjugate of

$$\Gamma(y) = \int_{\mathcal{Z}} \gamma(\langle y, \theta \rangle) dR, \quad y \in \mathcal{Y}_o.$$

Theorem 2.14 (Problem (P_C)). *Suppose that*

- (1) *the assumptions (A) and (A _{θ^{\vee}}) are satisfied;*
- (2) *C satisfies (2.9).*

Then:

(a) *The following dual equality for (P_C) holds:*

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

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

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

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

$$\begin{cases} (a) & \hat{x} \in C \cap \text{dom } \Gamma^* \\ (b) & \langle \tilde{\omega}, \hat{x} \rangle_{\mathcal{X}_o^*, \mathcal{X}_o} \leq \langle \tilde{\omega}, x \rangle_{\mathcal{X}_o^*, \mathcal{X}_o}, \forall x \in C \cap \text{dom } \Gamma^* \\ (c) & \hat{Q}(dz) = \gamma'_z(\langle \tilde{\omega}, \theta(z) \rangle) R(dz). \end{cases} \quad (2.15)$$

Furthermore, $\hat{Q} \in L_{\lambda_o^*} R$ and $\tilde{\omega} \in \tilde{\mathcal{Y}}$ satisfy (2.15) if and only if \hat{Q} solves (P_C) and $\tilde{\omega}$ solves (\bar{D}_C) .

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

$$\hat{x} = \int_{\mathcal{Z}} \theta \gamma'(\langle \tilde{\omega}, \theta \rangle) dR \quad (2.16)$$

in the weak sense. Moreover,

1. \hat{x} minimizes Γ^* on C ,
2. $I(\hat{Q}) = \Gamma^*(\hat{x}) = \int_{\mathcal{Z}} \gamma^* \circ \gamma'(\langle \tilde{\omega}, \theta \rangle) dR < \infty$ and
3. $I(\hat{Q}) + \int_{\mathcal{Z}} \gamma(\langle \tilde{\omega}, \theta \rangle) dR = \int_{\mathcal{Z}} \langle \tilde{\omega}, \theta \rangle d\hat{Q}$.

Proof. This result is [15, Theorem 5.2]. □

Following the terminology of Ney [22, 23], as it shares the properties (a), (b) and (2.16), the minimizer \hat{x} is called a *dominating point* of C with respect to Γ^* (see Definition 6.2 below).

2.8. The solutions of (\bar{P}_C) . Let us turn our attention to the extended problem (\bar{P}_C) . Denoting the operator $T : \ell \in L'_{\lambda_o} \mapsto \langle \theta, \ell \rangle \in \mathcal{X}_o$, it is shown in [15] that $TL'_{\lambda_o} \subset \mathcal{X}_L$ where $\mathcal{X}_L \subset \mathcal{X}_o$ is the topological dual space of $(\mathcal{Y}_L, |\cdot|_L)$ with $|y|_L = \|\langle y, \theta(\cdot) \rangle\|_{L_{\lambda_o}}$. Hence $T : L'_{\lambda_o} \rightarrow \mathcal{X}_L$ and one can define its adjoint $T^* : \mathcal{X}_L^* \rightarrow L'_{\lambda_o}$ for all $\omega \in \mathcal{X}_L^*$ by: $\langle \ell, T^*\omega \rangle_{L'_{\lambda_o}, L'_{\lambda_o}} = \langle T\ell, \omega \rangle_{\mathcal{X}_L, \mathcal{X}_L^*}, \forall \ell \in L'_{\lambda_o}$. We have the inclusions $\mathcal{Y}_o \subset \mathcal{Y}_L \subset \mathcal{X}_L^*$. The extended dual problem is

$$\text{maximize } \inf_{x \in C \cap \mathcal{X}_L} \langle \omega, x \rangle - I_\gamma([T^*\omega]_1^a) - \iota_D([T^*\omega]_2), \quad \omega \in \bar{\mathcal{Y}} \quad (\bar{D}_C)$$

where D is the $\sigma(K_\lambda^{s'}, K_\lambda^s)$ -closure of $\text{dom } I_\gamma$ and $\bar{\mathcal{Y}}$ is the cone of all $\omega \in \mathcal{X}_L^*$ such that $T^*\omega \in K_\lambda''$ and $[T^*\omega]_1^s = 0$.

As R is assumed to be σ -finite, there exists a measurable partition $(\mathcal{Z}_k)_{k \geq 1}$ of $\mathcal{Z} : \bigsqcup_k \mathcal{Z}_k = \mathcal{Z}$, such that $R(\mathcal{Z}_k) < \infty$ for each $k \geq 1$.

Theorem 2.17 (Problem (\bar{P}_C)). *Suppose that*

- (1) the assumptions (A) are satisfied;
- (2) C satisfies (2.10);
- (3) for each $k \geq 1$, $L_{\lambda_o}(\mathcal{Z}_k, R|_{\mathcal{Z}_k})$ is dense in $L_{\lambda_+}(\mathcal{Z}_k, R|_{\mathcal{Z}_k})$ and $L_{\lambda_-}(\mathcal{Z}_k, R|_{\mathcal{Z}_k})$ with respect to the topologies associated with $\|\cdot\|_{\lambda_+}$ and $\|\cdot\|_{\lambda_-}$.

Then:

(a) The following dual equality for (\bar{P}_C) holds:

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

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

Suppose that in addition we have

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

or equivalently $C \cap \text{icor}(T\text{dom } \bar{I}) \neq \emptyset$. Then:

(c) Let $\hat{\ell}$ be a solution to (\bar{P}_C) and denote $\hat{x} \triangleq T\hat{\ell}$. There exists $\bar{\omega} \in \bar{\mathcal{Y}}$ such that

$$\begin{cases} (a) & \hat{x} \in C \cap \text{dom } \Gamma^* \\ (b) & \langle \bar{\omega}, \hat{x} \rangle_{\mathcal{X}_L^*, \mathcal{X}_L} \leq \langle \bar{\omega}, x \rangle_{\mathcal{X}_L^*, \mathcal{X}_L}, \forall x \in C \cap \text{dom } \Gamma^* \\ (c) & \hat{\ell} \in \gamma'_z([T^*\bar{\omega}]_1^a) R + D^\perp([T^*\bar{\omega}]_2) \end{cases} \quad (2.18)$$

where

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

is the outer normal cone of D at η .

$T^*\bar{\omega}$ is in the $\sigma(K_\lambda'', K_\lambda')$ -closure of $T^*(\text{dom } \Gamma)$ and there exists some $\tilde{\omega} \in \mathcal{X}_\circ^*$ such that

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

is a measurable function in the strong closure of $T^*(\text{dom } \Gamma)$ in K_λ .

Furthermore, $\hat{\ell} \in L_{\lambda_\circ}'$ and $\bar{\omega} \in \bar{\mathcal{Y}}$ satisfy (2.18) if and only if $\hat{\ell}$ solves (\bar{P}_C) and $\bar{\omega}$ solves (\bar{D}_C) .

(d) Of course, (2.18-c) implies

$$\hat{x} = \int_{\mathcal{Z}} \theta \gamma'(\langle \tilde{\omega}, \theta \rangle) dR + \langle \theta, \hat{\ell}^s \rangle. \quad (2.19)$$

Moreover,

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

Proof. This result is [15, Theorem 4.6] which is proved without assuming $(A_{\gamma^*}^4)$. \square

Proposition 2.20. For the assumption (3) of Theorem 2.17 to be satisfied, it is enough that one of these conditions holds

- (i) λ is even or more generally $0 < \liminf_{t \rightarrow \infty} \frac{\lambda_+}{\lambda_-}(t) \leq \limsup_{t \rightarrow \infty} \frac{\lambda_+}{\lambda_-}(t) < +\infty$;
- (ii) $\lim_{t \rightarrow \infty} \frac{\lambda_+}{\lambda_-}(t) = +\infty$ and λ_- satisfies the Δ_2 -condition (2.4).

Proof. This result is [15, Proposition 4.9]. \square

Remark 2.21. Since it is assumed that $x \in \text{icordom } \Gamma^*$, no infinite force field (see [18] for this notion) enters the dual representation of ℓ_x . If 0 is the left bound of $\text{dom } \gamma^*$, one has $\gamma'(-\infty) = 0$ and in case $x \in \text{dom } \Gamma^* \setminus \text{icordom } \Gamma^*$, it may happen that $\langle y_x, \theta \rangle$ takes (formally) the value $-\infty$ on some subset \mathcal{Z}_0 of \mathcal{Z} so that $\frac{d\ell_x^a}{dR}$ vanishes on \mathcal{Z}_0 . A similar phenomenon occurs if $\text{dom } \gamma_z^*$ admits a finite right bound κ_z on a set \mathcal{Z}_κ such that $R(\mathcal{Z}_\kappa) > 0$. Then, $\gamma'_z(+\infty) = \kappa_z < \infty$ and $\frac{d\ell_x^a}{dR}(z) = \kappa_z$ on \mathcal{Z}_κ . For the details, see [18].

3. SOME EXAMPLES

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

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

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

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

$$\lambda_z(s) = e^s - s - 1, \quad s \in \mathbb{R}, z \in \mathcal{Z}. \quad (3.1)$$

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

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

This entropy is the rate function of (1.3) when $(W_i)_{i \geq 1}$ is an iid sequence of Poisson(1) random weights. If R is σ -finite, it is the rate function of the LDP of normalized Poisson random measures, see [16].

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

$$\begin{aligned} L_{\tau^*} &= \left\{ f : \mathcal{Z} \rightarrow \mathbb{R}; \int_{\mathcal{Z}} |f| \log |f| dR < \infty \right\} \\ E_{\tau} &= \left\{ u : \mathcal{Z} \rightarrow \mathbb{R}; \forall \alpha > 0, \int_{\mathcal{Z}} e^{\alpha|u|} dR < \infty \right\} \\ L_{\tau} &= \left\{ u : \mathcal{Z} \rightarrow \mathbb{R}; \exists \alpha > 0, \int_{\mathcal{Z}} e^{\alpha|u|} dR < \infty \right\} \end{aligned}$$

The extended relative entropy is defined by

$$\bar{I}(\ell|R) = I(\ell^a|R) + \sup \left\{ \langle \ell^s, u \rangle; u, \int_{\mathcal{Z}} e^u dR < \infty \right\}, \quad \ell \in \mathcal{E}(\mathcal{Z}) \quad (3.2)$$

where $\ell = \ell^a + \ell^s$ is the decomposition into absolutely continuous and singular parts of ℓ in $L'_{\tau} = L_{\tau^*} \oplus L_{\tau}^s$, and $\mathcal{E}(\mathcal{Z}) = \{\ell \in L'_{\tau}; \ell \geq 0, \langle \ell, \mathbf{1} \rangle = 1\}$. Note that $\mathcal{E}(\mathcal{Z})$ depends on R and that for all $\ell \in \mathcal{E}(\mathcal{Z})$, $\ell^a \in P_{\mathcal{Z}} \cap L_{\tau^*} R$.

3.2. Some examples of constraints. Let us consider the two standard constraints which are the moment constraints and the marginal constraints.

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

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

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

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

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

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

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

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

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

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

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

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

$$\langle (f, g), \theta \rangle = f \oplus g \in U_{AB}, \quad f \in U_A, g \in U_B \quad (3.3)$$

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

4. SOLVING (P_C) WITH BAD CONSTRAINTS

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

4.1. The results about minimizing sequences. One starts this section stating its main results at Theorems 4.1 and 4.7.

Theorem 4.1. *The hypotheses of Theorem 2.17 are assumed.*

- (a) *Suppose that $C \cap \text{dom } \Gamma^* \neq \emptyset$. Then, the minimization problem (\overline{P}_C) is attained in L'_{λ_\circ} and all its solutions share the same unique absolutely continuous part $Q_\circ \in L_{\lambda_\circ^*} R$.*
 (b) *Suppose that $C \cap \text{icordom } \Gamma^* \neq \emptyset$. Then, (\widetilde{D}_C) is attained in $\widetilde{\mathcal{Y}}$ and*

$$Q_\circ(dz) = \gamma'_z(\langle \omega_\circ, \theta(z) \rangle) R(dz) \quad (4.2)$$

where $\omega_\circ \in \widetilde{\mathcal{Y}}$ is any solution to (\widetilde{D}_C) . More, for each such ω_\circ , $\langle \omega_\circ, \theta(\cdot) \rangle$ is in the strong closure of $T^* \text{dom } \Gamma$ in K_λ and there exists $\bar{\omega} \in \overline{\mathcal{Y}}$ solution of (\overline{D}_C) such that $[T^* \bar{\omega}]_1^a = \langle \omega_\circ, \theta(\cdot) \rangle$.

Proof. • Proof of (a). The attainment statement is Theorem 2.17-b. Let us show that as γ^* is strictly convex, if k_* and ℓ_* are two solutions of (\overline{P}_C) , their absolutely continuous parts match:

$$k_*^a = \ell_*^a. \quad (4.3)$$

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

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

It follows that $I(pk_*^a + q\ell_*^a) + I^s(pk_*^s + q\ell_*^s) = pI(k_*^a) + qI(\ell_*^a) + pI^s(k_*^s) + qI^s(\ell_*^s)$. Suppose that $k_*^a \neq \ell_*^a$. As I is strictly convex, with $0 < p < 1$, 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 (4.3).

• Proof of (b). Let $\bar{\ell}$ be any solution to (\bar{P}_C) . Denoting $\bar{x}^a = T\bar{\ell}^a$ and $\bar{x}^s = T\bar{\ell}^s$ one sees with (2.18) that

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

By Theorem 2.17-c, this implies that $\tilde{\omega}$ solves $(\tilde{D}_{C-\bar{x}^s})$.

It remains to show that $\tilde{\omega}$ also solves (\tilde{D}_C) . Thanks to Theorem 2.17, we have: $\inf_{x \in C} \langle \bar{\omega}, x \rangle = \langle T^*\bar{\omega}, \bar{\ell} \rangle = \langle \langle \bar{\omega}, \theta \rangle, \bar{\ell}^a \rangle + \bar{I}(\bar{\ell}^s)$ and $\inf_{x \in C} \langle \bar{\omega}, x \rangle - I_\gamma(\langle \bar{\omega}, \theta \rangle) = \sup(\bar{D}_C) = \inf(\bar{P}_C) = I(\bar{\ell}^a) + \bar{I}(\bar{\ell}^s) = \inf(\bar{P}_{C-\bar{x}^s}) + \bar{I}(\bar{\ell}^s) = \sup(\bar{D}_{C-\bar{x}^s}) + \bar{I}(\bar{\ell}^s)$. Therefore $\inf_{x \in C} \langle \bar{\omega}, x \rangle = \inf_{x \in C-\bar{x}^s} \langle \bar{\omega}, x \rangle + \bar{I}(\bar{\ell}^s)$ and subtracting $\bar{I}(\bar{\ell}^s)$ from (\bar{D}_C) , one sees that $\tilde{\omega}$ which solves $(\tilde{D}_{C-\bar{x}^s})$ also solves (\tilde{D}_C) . One completes the proof of the proposition, taking $\omega_\diamond = \tilde{\omega}$. \square

Remark 4.4. Replacing $\bar{\ell}^s$ with $t\bar{\ell}^s$, the same proof shows that $\tilde{\omega}$ solves $(\tilde{D}_{C+(t-1)\bar{x}^s})$ for any $t \geq 0$.

From now on, one denotes $Q_\diamond \in L_{\lambda_\diamond} R$ the absolutely continuous part shared by all the solutions of (\bar{P}_C) . Let us introduce

$$\begin{aligned} \mathcal{C} &= \left\{ Q \in L_{\lambda_\diamond} R; TQ := \int_{\mathcal{Z}} \theta dQ \in C \right\} \\ \bar{\mathcal{C}} &= \left\{ \ell \in L'_{\lambda_\diamond}; T\ell := \langle \theta, \ell \rangle \in C \right\} \end{aligned} \quad (4.5)$$

the constraint sets $T^{-1}C \cap L_{\lambda_\diamond} R$ and $T^{-1}C$ on which I and \bar{I} are minimized. We have: $\mathcal{C} = \bar{\mathcal{C}} \cap L_{\lambda_\diamond} R$ and $I = \bar{I} + \iota_{L_{\lambda_\diamond} R}$. Hence, $\inf(\bar{P}_C) \leq \inf(P_C)$. Clearly, $\mathcal{C} \cap \text{dom } I \neq \emptyset \Leftrightarrow \inf(P_C) < \infty$ implies $\bar{\mathcal{C}} \cap \text{dom } \bar{I} \neq \emptyset \Leftrightarrow \inf(\bar{P}_C) < \infty \Leftrightarrow C \cap \text{dom } \Gamma^* \neq \emptyset$. Of course, if $\mathcal{C} \cap \text{dom } I \neq \emptyset$, (P_C) admits nontrivial minimizing sequences. Theorem 4.7 below gives some details about them.

The present paper is concerned with sets \mathcal{C} of the form (4.5), but this is not a restriction as explained in the following remarks.

Remarks 4.6.

- (1) Taking T^* to be the identity on $\mathcal{Y}_\diamond = E_{\lambda_\diamond}$ or L_{λ_\diamond} (being careless with a.e. equality, this corresponds to $\theta(z)$ to be the Dirac measure δ_z), one sees that (A_θ^\vee) or (A_θ^\exists) is satisfied respectively. Hence, with $C = \mathcal{C}$, (P_C) is (\mathcal{P}_C) .

Consequently, the specific form with θ and C adds details to the description of \mathcal{C} without loss of generality.

- (2) With $\theta, \mathcal{Y}_\diamond$ and C as in (1), the assumption (A_C) is:

- (a) Under (A_θ^\forall) : \mathcal{C} is $\sigma(L_{\lambda_\diamond^*}R, E_{\lambda_\diamond})$ -closed.
 Note that if λ_\diamond and λ_\diamond^* both satisfy the Δ_2 -condition, as \mathcal{C} is convex, this is equivalent to \mathcal{C} is $\|\cdot\|_{\lambda_\diamond^*}$ -closed.
- (b) Under (A_θ^\exists) : \mathcal{C} is $\sigma(L_{\lambda_\diamond^*}R, L_{\lambda_\diamond})$ -closed.
 Note that if λ_\diamond^* satisfies the Δ_2 -condition, as \mathcal{C} is convex, this is equivalent to \mathcal{C} is $\|\cdot\|_{\lambda_\diamond^*}$ -closed.

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

Theorem 4.7 (Minimizing sequences of (P_C)). *Suppose that the assumptions of Theorem 2.17 are satisfied and $\mathcal{C} \cap \text{dom } I \neq \emptyset$.*

Let us consider the following additional conditions.

- (1) a- *There are finitely many moment constraints, i.e. $\mathcal{X}_o = \mathbb{R}^K$ (see Section 3.2.1)*
 b- *$\mathcal{C} \cap \text{icordom } I \neq \emptyset$.*
- (2) *\mathcal{C} is a $\sigma(L_{\lambda_\diamond^*}R, L_{\lambda_\diamond})$ -closed convex set with a nonempty $\|\cdot\|_{\lambda_\diamond^*}$ -interior.*

Under one of these additional conditions (1) or (2), any minimizing sequence of (P_C) converges to Q_\diamond with respect to $\sigma(L_{\lambda_\diamond^}R, E_{\lambda_\diamond})$ and $I(Q_\diamond) \leq \inf(P_C) = \inf(\overline{P}_C)$.*

Proof. This proof relies on results which are stated and proved in the remainder of the present section. It is shown at Lemma 4.9 that any minimizing sequence of (P_C) converges in the sense of the $\sigma(L_{\lambda_\diamond^*}R, E_{\lambda_\diamond})$ -topology to Q_\diamond , whenever $\inf(P_C) = \inf(\overline{P}_C)$. But this equality holds thanks to Lemma 4.19 and

- under condition (1): Lemma 4.21-a;
- under condition (2): Corollary 4.23-b.

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

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

with a strict inequality if $I^s(\bar{\ell}^s) > 0$. □

Remarks 4.8.

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

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

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

- (ii) For \mathcal{C} to have a nonempty $\|\cdot\|_{\lambda_\diamond^*}$ -interior, it is enough that $\mathcal{C} \cap \mathcal{X}_L$ has a nonempty interior in \mathcal{X}_L endowed with the uniform dual norm $|\cdot|_L^*$. This is a consequence of [14, Lemma 4.13-e].
- (3) The last quantity $I^s(\bar{\ell}^s) = \inf(P_C) - I(Q_\diamond)$ is precisely the *gap* of lower $\sigma(L_{\lambda_\diamond^*}R, E_{\lambda_\diamond})$ -semicontinuity of $I : \lim_n Q_n = Q_\diamond$ and $\inf(P_C) = \lim \inf_n I(Q_n) \geq I(\lim_n Q_n) = I(Q_\diamond)$.

4.2. A preliminary lemma. Preliminary results for the proof of Theorem 4.7 are stated below at Lemma 4.9. The assumption (A_θ^\exists) about the bad constraint is $T^*\mathcal{Y}_o \subset L_{\lambda_o}$.

Lemma 4.9. *Suppose that the assumptions of Theorem 2.17 are satisfied, $\inf(P_C) < \infty$ and*

$$\inf(P_C) = \inf(\overline{P}_C) \quad (4.10)$$

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

Proof. Let $(Q_n)_{n \geq 1}$ be a minimizing sequence of (P_C) . Since it is assumed that $\inf(P_C) = \inf(\overline{P}_C)$, $(Q_n)_{n \geq 1}$ is also a minimizing sequence of (\overline{P}_C) . With the representation of \overline{I} at (4.11) below, one can apply ([17], Corollary 2.2) which states that \overline{I} is $\sigma(L'_{\lambda_o}, L_{\lambda_o})$ -inf-compact. Hence, one can extract a $\sigma(L'_{\lambda_o}, L_{\lambda_o})$ -convergent subnet $(Q_\alpha)_\alpha$ from $(Q_n)_{n \geq 1}$. Let $\ell_* \in \overline{C}$ denote its limit: we have $\lim_\alpha \int_{\mathcal{Z}} u dQ_\alpha = \langle \ell_*, u \rangle$ for all $u \in L_{\lambda_o}$. As $\langle \ell_*^s, u \rangle = 0$, for all $u \in E_{\lambda_o}$ (see Theorem 2.2-a), we obtain: $\lim_\alpha \int_{\mathcal{Z}} u dQ_\alpha = \int_{\mathcal{Z}} u d\ell_*^a$ for all $u \in E_{\lambda_o}$. This proves that $(Q_\alpha)_\alpha$ $\sigma(E'_{\lambda_o}, E_{\lambda_o})$ -converges to ℓ_*^a . As E_{λ_o} is a separable Banach space (L_{λ_o} is not separable in general), the topology $\sigma(E'_{\lambda_o}, E_{\lambda_o}) = \sigma(L_{\lambda_o^*}R, E_{\lambda_o})$ is metrizable and one can extract a convergent subsequence $(\tilde{Q}_k)_{k \geq 1}$ from the convergent net $(Q_\alpha)_\alpha$. Hence, $(\tilde{Q}_k)_{k \geq 1}$ is a subsequence of $(Q_n)_{n \geq 1}$ which $\sigma(L_{\lambda_o^*}R, E_{\lambda_o})$ -converges to ℓ_*^a . Since \overline{I} is inf-compact, ℓ_* is a minimizer of (\overline{P}_C) and by Theorem 4.1-a, there is a unique Q_\diamond such for any minimizing sequence $(Q_n)_{n \geq 1}$, $\ell_*^a = Q_\diamond$. Therefore, any convergent subsequence of $(Q_n)_{n \geq 1}$ converges to Q_\diamond for $\sigma(L_{\lambda_o^*}R, E_{\lambda_o})$. As any subsequence of a minimizing sequence is still a minimizing sequence, we have proved that from any subsequence of $(Q_n)_{n \geq 1}$, one can extract a sub-subsequence which converges to Q_\diamond . This proves that $(Q_n)_{n \geq 1}$ converges to Q_\diamond with respect to $\sigma(L_{\lambda_o^*}R, E_{\lambda_o})$. \square

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

$$\begin{aligned} \Phi_L(u) &= I_\lambda(u) = \int_{\mathcal{Z}} \lambda(u) dR, \quad u \in L_{\lambda_o} \\ \Phi_E(u) &= \Phi_L(u) + \iota_E(u), \quad u \in L_{\lambda_o} \end{aligned}$$

Their convex conjugates are

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

It is shown in [17] that

$$\begin{cases} I(\ell) = \Phi_E^*(\ell - mR), & \ell \in E'_{\lambda_o} = L_{\lambda_o^*}R \\ \overline{I}(\ell) = \Phi_L^*(\ell - mR), & \ell \in L'_{\lambda_o} = L_{\lambda_o^*}R \oplus L_{\lambda_o}^s \end{cases} \quad (4.11)$$

Hence, considering the minimization problems

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

and

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

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

$$\inf(P_E) = \inf(P_L) \quad (4.12)$$

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

Basic facts about convex duality. The proof of (4.12) will rely on standard convex duality considerations. Let us recall some related facts, as developed in [26].

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

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

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

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

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

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

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

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

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

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

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

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

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

Theorem 4.13 (Criteria for the dual equality). *We assume that A, U, X and Y are locally convex Hausdorff topological vector spaces such that $\langle A, U \rangle$ and $\langle X, Y \rangle$ are topological dual pairings. For the dual equality*

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

to hold, it is enough that

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

or

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

Remarks 4.14. About the space U .

- (a) As regards Criterion (1), the space U is unnecessary.
- (b) As regards Criterion (2), it is not assumed that $T^*Y \subset U$.

Back to our problem. Let us particularize this framework for the problems (P_E) and (P_L) . Assuming that $m \equiv 0$, one sees that $(P_C) = (P_E)$, $(\bar{P}_C) = (P_L)$, $I = I_{\lambda^*} = \Phi_E^*$, $\bar{I} = \bar{I}_{\lambda^*} = \Phi_L^*$, $\gamma = \lambda$, $C = C_o$ and so on. *This simplifying requirement will be assumed during the proof without loss of generality, see the proof of [15, Theorem 4.6].*

Let us first apply the criterion (1) of Theorem 4.13.

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

$$\varphi_E(x) = \inf \left\{ I(Q); \int_{\mathcal{Z}} \theta dQ \in C - x, Q \in L_{\lambda^*}R \right\}, \quad x \in \mathcal{X}_L$$

Under the underlying assumption (A_{θ}^{\exists}) , with [14, Lemma 4.13-g] we have:

$$T^*\mathcal{Y}_L \subset L_{\lambda_o}. \quad (4.15)$$

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

$$h^*(u) = \int_{\mathcal{Z}} \lambda(u) dR, \quad u \in L_{\lambda_o} \quad (4.16)$$

Therefore, the dual problem associated to (P_E) is

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

Let us go on with (P_L) . Take $A = L'_{\lambda_o}$, $X = \mathcal{X}_L$, $Y = \mathcal{Y}_L$ equipped with the weak topologies $\sigma(\mathcal{X}_L, \mathcal{Y}_L)$ and $\sigma(\mathcal{Y}_L, \mathcal{X}_L)$, and $h = \Phi_L^* = \bar{I}$. The function h^* in restriction to L_{λ_o} is still given by (4.16) since $u \in L_{\lambda_o} \mapsto \int_{\mathcal{Z}} \lambda(u) dR$ is closed convex (Fatou's lemma). The primal value-function is

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

and the dual problem associated to (P_L) is

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

Lemma 4.17. *Suppose that $T^*\mathcal{Y}_L \subset L_{\lambda_o}$ and $C \cap \mathcal{X}_L$ is $\sigma(\mathcal{X}_L, \mathcal{Y}_L)$ -closed, then φ_L is $\sigma(\mathcal{X}_L, \mathcal{Y}_L)$ -lower semicontinuous.*

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

As already seen, \bar{I} is $\sigma(L'_{\lambda_o}, L_{\lambda_o})$ -inf-compact and T is $\sigma(L'_{\lambda_o}, L_{\lambda_o})$ - $\sigma(\mathcal{X}_L, \mathcal{Y}_L)$ -continuous, see [14, Lemma 4.13-h]. It follows that \bar{J} is $\sigma(\mathcal{X}_L, \mathcal{Y}_L)$ -inf-compact. As $C \cap \mathcal{X}_L$ is assumed to be $\sigma(\mathcal{X}_L, \mathcal{Y}_L)$ -closed, ι_{-C} is lower semicontinuous. Finally, being the inf-convolution of an inf-compact function and a lower semicontinuous function, $\tilde{\varphi}$ is lower semicontinuous, and so is φ_L . \square

As \bar{I} and C are assumed to be convex and $\sup(D_L) \geq \inf_{x \in C} \langle 0, x \rangle - \int_{\mathcal{Z}} \lambda(T^*0) dR = 0 > -\infty$, this lemma allows us to apply Criterion (1) of Theorem 4.13 to obtain

$$\inf(P_L) = \sup(D_L) \quad (4.18)$$

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

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

Since the desired equality (4.10) is equivalent to $\inf(P_L) = \inf(P_E)$, we have proved

Lemma 4.19. *The equality (4.10) holds if and only if we have the dual equality*

$$\inf(P_E) = \sup(D_E). \quad (4.20)$$

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

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

Lemma 4.21.

- (a) *Suppose that there are finitely many constraints (i.e. \mathcal{X}_o is finite dimensional) and $\mathcal{C} \cap \text{icordom } I \neq \emptyset$, then φ_E is continuous at 0.*
- (b) *Suppose that (A_θ^\vee) is satisfied and \mathcal{C} is $\sigma(\mathcal{X}_E, \mathcal{Y}_E)$ -closed, then φ_E is $\sigma(\mathcal{X}_L, \mathcal{Y}_L)$ -lower semicontinuous.*

Proof. • Proof of (a). To get (a), remark that a convex function on a finite dimensional space is lower semicontinuous on the intrinsic core of its effective domain. By [14, Lemma 4.13], T is $\sigma(L_{\lambda_\circ}^* R, L_{\lambda_\circ})$ -continuous and the assumption $\mathcal{C} \cap \text{icordom } I \neq \emptyset$ implies that 0 belongs to $\text{icordom } \varphi_E$.

• Proof of (b). It is similar to the proof of Lemma 4.17. The assumption (A_θ^\vee) insures that $\mathcal{X}_E = \mathcal{X}_L$ and T is $\sigma(L_{\lambda_\circ}^* R, E_{\lambda_\circ})$ - $\sigma(\mathcal{X}_E, \mathcal{Y}_E)$ -continuous. \square

It follows from Lemma 4.21-b, Lemma 4.19, the remark at (4.12) and Lemma 4.9 that under the good constraint assumptions (A^\vee) , if $\mathcal{C} \cap \text{dom } \Gamma^* \neq \emptyset$, then any minimizing sequence of (P_C) converges with respect to the topology $\sigma(L_{\lambda_\circ}^* R, E_{\lambda_\circ})$ to the unique solution \widehat{Q} of (P_C) . This is Theorem 2.14-b.

Using Criterion (2). Up to now, we only used Criterion (1) of Theorem 4.13. In the following lines, we are going to use Criterion (2) to prove (4.20) under additional assumptions.

Let us go back to Problem (P_E) . It is still assumed without loss of generality that $m = 0$ and $\gamma = \lambda$. One introduces a space U and a dual value-function ψ on U . The framework of Theorem 4.13 is preserved when taking $X = \mathcal{X}_L$, $Y = \mathcal{Y}_L$ with the weak topologies $\sigma(\mathcal{X}_L, \mathcal{Y}_L)$ and $\sigma(\mathcal{Y}_L, \mathcal{X}_L)$, $h = I$ on $A = L_{\lambda_\circ}^* R$ as before and adding the following topological pairing $\langle A, U \rangle$. We endow $A = L_{\lambda_\circ}^* R$ with the topology $\sigma(L_{\lambda_\circ}^* R, L_{\lambda_\circ})$ and take

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

with the topology $\sigma(L_{\lambda_\circ}, L_{\lambda_\circ}^*)$. By (4.16): $h^* = I_\lambda$, this leads to the dual value-function

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

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

Lemma 4.22. *For ψ to be $\sigma(L_{\lambda_\circ}, L_{\lambda_\circ}^*)$ -upper semicontinuous, it is enough that*

- (a) *$T^* \mathcal{Y}_L$ is a $\sigma(L_{\lambda_\circ}, L_{\lambda_\circ}^*)$ -closed subspace of L_{λ_\circ} and*
- (b) *the interior of \mathcal{C} in $L_{\lambda_\circ}^* R$ with respect to $\|\cdot\|_{\lambda_\circ^*}$ is nonempty.*

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

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

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

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

As V is assumed to be closed, we have $\iota_V = \iota_{V^\perp}^*$ with $V^\perp = \{k \in L_{\lambda_\circ}^* R; \langle v, k \rangle = 0, \forall v \in V\}$. This gives for each $u \in L_{\lambda_\circ}^*$, $G(u) = \iota_{\mathcal{C}}^*(u) + \iota_{V^\perp}^*(u) = \sup_{\ell \in C} \langle u, \ell \rangle + \sup_{k \in V^\perp} \langle u, k \rangle = \sup\{\langle u, \ell + k \rangle; \ell \in C, k \in V^\perp\} = \iota_{\mathcal{C} + V^\perp}^*(u) = \iota_{\mathcal{C}}^*(u)$, where the last equality holds since $\mathcal{C} + V^\perp = \mathcal{C}$, note that $\ker T = V^\perp$.

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

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

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

we have

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

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

Corollary 4.23.

- (a) Assume that C is $\sigma(\mathcal{X}_L, \mathcal{Y}_L)$ -closed convex, $T^*\mathcal{Y}_L$ is a $\sigma(L_{\lambda_\circ}, L_{\lambda_\circ}^*)$ -closed subspace of L_{λ_\circ} , \mathcal{C} has a nonempty $\|\cdot\|_{\lambda_\circ^*}$ -interior and $\inf(\mathbf{P}_E) < \infty$. Then, (4.20) is satisfied.
(b) In particular, if \mathcal{C} is $\sigma(L_{\lambda_\circ}^* R, L_{\lambda_\circ})$ -closed convex set with a nonempty $\|\cdot\|_{\lambda_\circ^*}$ -interior and $\inf(\mathbf{P}_E) < \infty$, then (4.20) is satisfied.

Proof. • Proof of (a). Apply the criterion (2) of Theorem 4.13 with Lemma 4.22.

- Proof of (b). This is (a) with $\mathcal{Y}_L = L_{\lambda_\circ}$ and $T^* = \text{Id}$, taking advantage of Remarks 4.6. \square

5. ENTROPIC PROJECTIONS

The results of the preceding sections are translated in terms of entropic projections.

5.1. Generalized entropic projections. We consider the convex problem

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

where \mathcal{C} is a convex subset of $M_{\mathcal{Z}}$.

Let $(f_n)_{n \geq 1}$ be a sequence of measurable functions. Following [4], one says that it converges R -loosely in measure to f if for each measurable set $A \in \mathcal{Z}$ such that $R(A) < \infty$ and all $\epsilon > 0$, $\lim_{n \rightarrow \infty} R(A \cap \{|f_n - f| > \epsilon\}) = 0$.

It converges $\sigma(L_{\lambda_\circ}^* R, E_{\lambda_\circ})$ -loosely if for each measurable set $A \in \mathcal{Z}$ such that $R(A) < \infty$, $\mathbf{1}_A f_n$ converges to $\mathbf{1}_A f$ with respect to $\sigma(L_{\lambda_\circ}^* R, E_{\lambda_\circ})$.

It converges L_1 -loosely if for each measurable set $A \in \mathcal{Z}$ such that $R(A) < \infty$, $\mathbf{1}_A f_n$ converges to $\mathbf{1}_A f$ strongly in $L_1(R)$.

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

- (1) One says that $(Q_n)_{n \geq 1}$ converges R -loosely in measure to Q if $\frac{dQ_n}{dR}$ converges R -loosely in measure to $\frac{dQ}{dR}$.
- (2) One says that $(Q_n)_{n \geq 1}$ converges $\sigma(L_{\lambda^*}R, E_{\lambda^*})$ -loosely to Q in $L_{\lambda^*}R$ if $\frac{dQ_n}{dR}$ converges $\sigma(L_{\lambda^*}R, E_{\lambda^*})$ -loosely to $\frac{dQ}{dR}$.
- (3) One says that $(Q_n)_{n \geq 1}$ converges R -loosely in variation to Q if $\frac{dQ_n}{dR}$ converges L_1 -loosely to $\frac{dQ}{dR}$.

Definition 5.2 (Generalized entropic projection). [4, Csiszár]. Suppose that $\mathcal{C} \cap \text{dom } I \neq \emptyset$ and that any minimizing sequence of the problem $(\mathcal{P}_{\mathcal{C}})$ converges R -loosely in measure to some $Q_* \in M_{\mathcal{Z}}$.

This Q_* is called the generalized entropic projection of mR on \mathcal{C} with respect to I . It may not belong to \mathcal{C} . In case Q_* is in \mathcal{C} , it is called the entropic projection of mR on \mathcal{C} .

Theorem 5.3 (Csiszár, [4]). Suppose that (A_R) and (A_{γ^*}) are satisfied. Then, mR has a generalized projection on any convex subset \mathcal{C} of $M_{\mathcal{Z}}$ such that $\mathcal{C} \cap \text{dom } I \neq \emptyset$.

In [4] γ^* doesn't depend on the variable z , but the proof remains unchanged with a z -dependence.

Proposition 5.4. Let Q and $(Q_n)_{n \geq 1}$ in $M_{\mathcal{Z}}$ be absolutely continuous with respect to R such that $(Q_n)_{n \geq 1}$ is a minimizing sequence which converges loosely in R -measure to Q . Then,

- (a) $(Q_n)_{n \geq 1}$ converges R -loosely in variation to Q ;
- (b) $(Q_n)_{n \geq 1}$ converges $\sigma(L_{\lambda^*}R, E_{\lambda^*})$ -loosely to Q .

Proof. • Proof of (a). Because of de la Vallée Poussin's theorem (using assumption $(A_{\gamma^*}^4)$), any minimizing sequence $(Q_n)_{n \geq 1}$ is such that $\{\mathbf{1}_A \frac{dQ_n}{dR}, n \geq 1\}$ is uniformly integrable for each A such that $R(A) < \infty$. Therefore, if $(Q_n)_{n \geq 1}$ converges loosely in R -measure, it converges R -loosely in variation.

• Proof of (b). It is enough to consider a bounded measure R . If λ^* is a finite function, this follows from [24, Proposition 4.5.6]. If $\text{dom } \lambda^*$ is bounded, we have to show the $\sigma(L_{\infty}, L_1)$ -convergence of $f_n = \frac{dQ_n}{dR}$ to $f = \frac{dQ}{dR}$. But $(f_n)_{n \geq 1}$ is bounded in L_{∞} and the result follows. \square

As a direct consequence of Theorem 2.14, if the constraints are good, the generalized entropic projection is the entropic projection.

Proposition 5.5. Suppose that (A_R) and (A_{γ^*}) hold, \mathcal{C} is convex and $\sigma(L_{\lambda^*}R, E_{\lambda^*})$ -closed and $\mathcal{C} \cap \text{dom } I \neq \emptyset$. Then, the entropic projection Q_* exists and is equal to

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

where \widehat{Q} is the minimizer of $(\mathcal{P}_{\mathcal{C}})$ which is described at Theorem 2.14.

Proof. This is an easy corollary of Theorem 2.14. \square

A consequence of Proposition 5.5 is that for any $\sigma(L_{\lambda^*}R, E_{\lambda^*})$ -closed convex set \mathcal{C} , the generalized entropic projection is the entropic projection. This is essentially [4, Thm 3-(iii)].

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

Theorem 5.6. *Suppose that the assumptions of Theorem 2.17 are satisfied and \mathcal{C} given by (4.5) satisfies $\mathcal{C} \cap \text{dom } I \neq \emptyset$. Let us consider the additional conditions:*

- (1) a- *There are finitely many moment constraints, i.e. $\mathcal{X}_o = \mathbb{R}^K$ (see Section 3.2.1)*
 b- *$\mathcal{C} \cap \text{icordom } I \neq \emptyset$;*
- (2) *\mathcal{C} is a $\sigma(L_{\lambda_o^*}R, L_{\lambda_o})$ -closed convex set with a nonempty $\|\cdot\|_{\lambda_o^*}$ -interior.*

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

$$Q_* = Q_\diamond$$

the absolutely continuous component described at (4.2) and $I(Q_) \leq \inf_{\mathcal{C}} I$.*

Proof. This is a direct consequence of Theorems 5.3 and 4.7. □

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

$$\text{minimize } I(P|R) \text{ subject to } \int_{\mathcal{Z}} \theta dP \in C, \quad P \in P_{\mathcal{Z}} \quad (5.7)$$

and its extension is

$$\text{minimize } \bar{I}(\ell|R) \text{ subject to } \langle \theta, \ell \rangle \in C, \quad \ell \in \mathcal{E}(\mathcal{Z}) \quad (5.8)$$

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

$$\Xi(x) = \sup_{y \in \mathcal{Y}_o} \left\{ \langle y, x \rangle - \log \int_{\mathcal{Z}} e^{\langle y, \theta \rangle} dR \right\} \in [0, \infty], \quad x \in \mathcal{X}_o \quad (5.9)$$

Proposition 5.10 (Relative entropy subject to good constraints). *Let us assume that θ satisfies the “good constraint” assumption*

$$\forall y \in \mathcal{Y}_o, \int_{\mathcal{Z}} e^{\langle y, \theta(z) \rangle} R(dz) < \infty \quad (5.11)$$

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

(a) *The following dual equality holds:*

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

(b) *Suppose that in addition $C \cap \text{dom } \Xi \neq \emptyset$. Then, the minimization problem (5.7) has a unique solution \hat{P} in $P_{\mathcal{Z}}$, \hat{P} is the entropic projection of R on $C = \{P \in P_{\mathcal{Z}}, \int_{\mathcal{Z}} \theta dP \in C\}$.*

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

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

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

$$\hat{x} = \int_{\mathcal{Z}} \theta(z) \exp \left(\langle \tilde{\omega}, \theta(z) \rangle - \log \int_{\mathcal{Z}} e^{\langle \tilde{\omega}, \theta \rangle} dR \right) R(dz) \quad (5.12)$$

in the weak sense.

Proposition 5.13 (Relative entropy subject to bad constraints). *Let us assume that θ satisfies the “bad constraint” assumption*

$$\forall y \in \mathcal{Y}_o, \exists \alpha > 0, \int_{\mathcal{Z}} e^{\alpha |\langle y, \theta(z) \rangle|} R(dz) < \infty \quad (5.14)$$

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

(a) The following dual equality holds:

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

(b) Suppose that in addition $C \cap \text{dom } \Xi \neq \emptyset$. Then, the minimization problem (5.8) is attained in $\mathcal{E}(\mathcal{Z})$: the set of minimizers is nonempty, convex and $\sigma(L'_\tau, L_\tau)$ -compact. Moreover, all the minimizers share the same unique absolutely continuous part $P_\diamond \in P_{\mathcal{Z}} \cap L_{\tau^*} R$ which is the generalized entropic projection of R on C .

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

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

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

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

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

In Proposition 5.10, \hat{x} is the dominating point in the sense of Ney (see Definition 6.3) of C with respect to Ξ . The representation of \hat{x} has already been obtained for C with a nonempty topological interior in \mathbb{R}^d by Ney in [23] and in a Banach space setting by Einmahl and Kuelbs in [8]. The representation of the generalized projection P_\diamond is obtained with a very different proof by Csiszár [2] and ([3], Thm 3). Proposition 5.13 also extends corresponding results of Kuelbs [13] which are obtained in a Banach space setting.

For more details about the minimizers of (5.8), one can look at ([19], Theorem 3.4) where a characterization is obtained under the weakest assumption: $C \cap \text{dom } \Xi \neq \emptyset$.

Proof of Propositions 5.10 and 5.13. They are direct consequences of Theorems 2.14, 2.17, Proposition 5.5, Theorem 5.6 and Lemma 5.15 below. Note that λ which is given at (3.1) satisfies the condition (ii) of Proposition 2.20. \square

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

Lemma 5.15. *For all $x \in \mathcal{X}_o$,*

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

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

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

which is the desired result. \square

Example 5.16. Csiszár's example. Comparing Proposition 5.5 with Theorem 5.6, one may wonder if the $\sigma(L_{\lambda_*}^* R, E_{\lambda_o})$ -closedness of the convex set \mathcal{C} is critical for the existence of an entropic projection. The answer is affirmative. In [3, Example 3.2], Csiszár gives an interesting example where the generalized entropic projection can be explicitly computed in a situation where (P_C) is not attained, see also [7, Exercise 7.3.11]. This example is the following one.

Take the probability measure on $\mathcal{Z} = [0, \infty)$ defined by $R(dz) = a_0 \frac{e^{-z}}{1+z^3} dz$ where a_0 is the normalizing constant, I the relative entropy with respect to R , $\theta : z \in [0, \infty) \mapsto z \in \mathcal{X} = \mathbb{R}$ and $C = [c, \infty)$. This gives $\mathcal{C} = \{Q \in L_{\tau^*} R; \int_{[0, \infty)} z Q(dz) \geq c, Q([0, \infty)) = 1\}$. The point is that θ is in $L_{\tau}(R)$ but not in $E_{\tau}(R)$.

By Theorem 5.13, the generalized projection of R on \mathcal{C} exists and is equal to $P_y = a_y \frac{e^{(y-1)z}}{1+z^3} dz$ for some real $y \leq 1$, where a_y is the normalizing constant. If it belongs to \mathcal{C} , then $\int_{\mathcal{Z}} z P_y(dz) \geq c$. But $\sup_{y \leq 1} \int_{\mathcal{Z}} z P_y(dz) = \int_{\mathcal{Z}} z P_1(dz) = a_1 \int_{\mathcal{Z}} \frac{z}{1+z^3} dz := x_* < \infty$. Therefore, for any $c > x_*$, there are no entropic projection but only a generalized one.

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

More details about this example are given below at Example 6.10.

6. DOMINATING POINTS

The underlying assumptions are Theorem 2.17's ones. We are going to investigate some relations between dominating points and entropic projections. In the case where the constraint is good, Theorem 2.14 and Proposition 5.5 state that the generalized entropic projection is the entropic projection $Q_* = \widehat{Q}$, the minimizer \bar{x} is the dominating point of C , it is related to Q_* by the identity:

$$\bar{x} = \langle \theta, Q_* \rangle. \quad (6.1)$$

We now look at the situation where the constraint is bad. As remarked at Example 5.16, the above equality may fail. A necessary and sufficient condition (in terms of the function Γ^*) for \bar{x} to satisfy (6.1) is obtained at Theorem 6.8.

Following Ney [22, 23], let us introduce the following definition. A point $\hat{x} \in \mathcal{X}_o$ sharing the properties (a), (b) and (2.16) of Theorem 2.14 is called a *dominating point*.

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

- (a) $\hat{x} \in C \cap \text{dom } \Gamma^*$
- (b) *there exists some linear form $\tilde{\omega}$ on \mathcal{X}_o such that $\langle \tilde{\omega}, \hat{x} \rangle \leq \langle \tilde{\omega}, x \rangle$ for all $x \in C \cap \text{dom } \Gamma^*$,*

(c) $\langle \tilde{\omega}, \theta(\cdot) \rangle$ is measurable and $\hat{x} = \int_{\mathcal{Z}} \theta(z) \gamma'(z, \langle \tilde{\omega}, \theta(z) \rangle) R(dz)$.

In the special case where Γ^* is replaced by the Cramér transform Ξ defined at (5.9), taking Lemma 5.15 into account, this definition becomes the following one.

Definition 6.3. Let $C \subset \mathcal{X}_o$ be a convex set such that $C \cap \mathcal{X}_L$ is $\sigma(\mathcal{X}_L, \mathcal{Y}_L)$ -closed. The point $\hat{x} \in \mathcal{X}_o$ is called a Ξ -dominating point if

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

Note that this definition is slightly different from the ones proposed by Ney [23] and Einmahl and Kuelbs [8] since C is neither supposed to be an open set nor to have a non-empty interior and \hat{x} is not assumed to be a boundary point of C . The above integral representation (c) is (5.12).

Recall that the extended entropy \bar{I} is given by (2.7): $\bar{I}(\ell) = I(\ell^a) + I^s(\ell^s)$, $\ell = \ell^a + \ell^s \in L'_{\lambda_o}$ and define for all $x \in \mathcal{X}_o$,

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

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

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

where $J \square J^s$ is the inf-convolution of J and J^s . By Theorem 2.17-a, $\bar{J} = \Gamma^*$ and if $\bar{J}(x) < \infty$, there exists $\ell_x \in L'_{\lambda_o}$ such that $\langle \theta, \ell_x \rangle = x$ and $\bar{J}(x) = \bar{I}(\ell_x)$. Let us define

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

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

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

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

Proposition 6.4. For all $x \in \text{dom } \bar{J}$, we have:

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

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

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

Definition 6.5 (Recessive x). Let us say that x is recessive for Γ^* if for some $\delta > 0$ and $\xi \in \mathcal{X}_o$, $\Gamma^*(x + t\xi) - \Gamma^*(x) = t\widetilde{\Gamma^*}(\xi)$ for all $t \in (-\delta, +\infty)$. It is said to be non-recessive otherwise.

Proposition 6.6. *We have $\bar{J} = \Gamma^*$ and $J^s = \widetilde{\Gamma}^*$. Moreover, $J(x) = \Gamma^*(x)$ for all non-recessive $x \in \mathcal{X}_o$.*

Proof. We have already noted that $\bar{J} = \Gamma^*$ and by [14, Theorem 2.6-a], we get: $J^s = \iota_{\text{dom } \Gamma}^*$: the support function of $\text{dom } \Gamma$. Therefore, it is also the recession function of Γ^* . Hence, we have $\Gamma^* = \bar{J} = J \square J^s = J \square \widetilde{\Gamma}^*$.

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

Proposition 6.7. *For all $x \in \text{dom } \Gamma^*$, we have:*

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

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

Proof. By (2.7), we have $\bar{I}(\ell_x) = I(\ell_x^a) + \widetilde{I}(\ell_x^s)$ where \widetilde{I} is the recession function of I . It follows that $\bar{J}(x) = J(x^a) + J^s(x^s)$, since $J(x^a) = I(\ell_x^a)$ (Proposition 6.4) and the recession function of \bar{J} is J^s . To show this, note that (see [27])

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

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

Theorem 6.8. *Let us assume that the hypotheses of Theorem 2.17 hold and that $C \cap \text{icordom } \Gamma^* \neq \emptyset$.*

- (a) *Then, a minimizer \bar{x} of Γ^* on the set C is a Γ^* -dominating point of C if and only if \bar{x} is non-recessive. This is also equivalent to the following statement: “all the solutions of the minimization problem (\bar{P}_C) are absolutely continuous with respect to R .” In such a case the solution of (\bar{P}_C) is unique and it matches the solution of (P_C) .*
- (b) *In particular when Γ^* admits a degenerate recession function, i.e. $\widetilde{\Gamma}^*(x) = +\infty$ for all $x \neq 0$, then the minimizer \bar{x} is a Γ^* -dominating point of C .*
- (c) *The same statements hold with Γ^* replaced by Ξ .*

Proof. This is a direct consequence of Theorem 2.17, Proposition 6.7 and Lemma 5.15. \square

Remark 6.9. A remark about the steepness of the log-Laplace transform. In [13, Thm 1], with the setting of Section 5.2 where \mathcal{X}_o is a Banach space, Kuelbs proves a result that is slightly different from statement (b) of the above theorem. It is proved that the existence of a Ξ -dominating point for *all* convex sets C with a nonempty topological interior is equivalent to some property of the Gâteaux derivative of the log-Laplace transform $y \in \mathcal{X}'_o \mapsto \log \int_{\mathcal{X}_o} \exp(\langle y, x \rangle) R \circ \theta^{-1}(dx)$ 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.

Example 6.10. Csiszár’s example continued. We go on with Example 5.16. By Lemma 5.15, Γ^* is the Cramér transform of R (identifying $(1, x)$ with x). This means that Γ^* is the convex conjugate of the log-Laplace transform $\Gamma(y) = \log \int_{[0, \infty)} a \frac{e^{(y-1)z}}{1+z^3} dz$, $y \in \mathbb{R}$.

Clearly, $\text{dom } \Gamma = (-\infty, 1]$ and $\Gamma'(1^-) = \int_{[0, \infty)} z P_1(dz) := x_* < \infty$. It follows that for all $x \geq x_*$, $\Gamma^*(x) - \Gamma^*(x_*) = x - x_*$. One deduces from this that (x_*, ∞) is a set of recessive points. By Theorem 6.8, they cannot be dominating points. Note also that the log-Laplace transform Γ is not steep.

REFERENCES

- [1] C. Boucher, R.S. Ellis, and B. Turkington. Spatializing random measures: doubly indexed processes and the large deviation principle. *Ann. Probab.*, 27:297–324, 1999.
- [2] I. Csiszár. I -divergence geometry of probability distributions and minimization problems. *Annals of Probability*, 3:146–158, 1975.
- [3] I. Csiszár. Sanov property, generalized I -projection and a conditional limit theorem. *Annals of Probability*, 12:768–793, 1984.
- [4] I. Csiszár. Generalized projections for non-negative functions. *Acta Math. Hungar.*, 68:161–185, 1995.
- [5] I. Csiszár, F. Gamboa, and E. Gassiat. MEM pixel correlated solutions for generalized moment and interpolation problems. *IEEE Trans. Inform. Theory*, 45(7):2253–2270, 1999.
- [6] D. Dacunha-Castelle and F. Gamboa. Maximum d’entropie et problème des moments. *Ann. Inst. H. Poincaré. Probab. Statist.*, 26:567–596, 1990.
- [7] A. Dembo and O. Zeitouni. *Large Deviations Techniques and Applications. Second edition.* Applications of Mathematics 38. Springer Verlag, 1998.
- [8] U. Einmahl and J. Kuelbs. Dominating points and large deviations for random vectors. *Probab. Theory Relat. Fields*, 105:529–543, 1996.
- [9] R.S. Ellis, J. Gough, and J.V. Puli. The large deviations principle for measures with random weights. *Rev. Math. Phys.*, 5:659–692, 1993.
- [10] F. Gamboa and E. Gassiat. Bayesian methods and maximum entropy for ill-posed inverse problems. *Ann. Stat.*, 25(1):328–350, 1997.
- [11] A. Kozek. Orlicz spaces of functions with values in Banach spaces. *Ann. Soc. Math. Polonae, Series I, Commentationes Mathematicae XIX*:259–288, 1977.
- [12] A. Kozek. Convex integral functionals on Orlicz spaces. *Ann. Soc. Math. Polonae, Series I, Commentationes Mathematicae XXI*:109–135, 1979.
- [13] J. Kuelbs. Large deviation probabilities and dominating points for open convex sets: nonlogarithmic behavior. *Ann. Probab.*, 28(3):1259–1279, 2000.
- [14] C. Léonard. Convex minimization problems with weak constraint qualifications. Preprint.
- [15] C. Léonard. Minimization of entropy functionals. Preprint.
- [16] C. Léonard. Large deviations for Poisson random measures and processes with independent increments. *Stochastic Processes and their Applications*, 85:93–121, 2000.
- [17] C. Léonard. Convex conjugates of integral functionals. *Acta Math. Hungar.*, 93(4):253–280, 2001.
- [18] C. Léonard. Minimizers of energy functionals. *Acta Math. Hungar.*, 93(4):281–325, 2001.
- [19] C. Léonard. Minimizers of energy functionals under not very integrable constraints. *Journal of Convex Analysis*, 10(1):63–88, 2003.
- [20] C. Léonard and J. Najim. An extension of Sanov’s theorem : application to the Gibbs conditioning principle. *Bernoulli*, 8(6):721–743, 2002.
- [21] J. Najim. A Cramér type theorem for weighted random variables. *Electronic Journal of Probability*, 7(4):1–32, 2002.
- [22] P. Ney. Dominating points and the asymptotics of large deviations for random walks on R^d . *Ann. Probab.*, 11:158–167, 1983.
- [23] P. Ney. Convexity and large deviations. *Ann. Probab.*, 12:903–906, 1984.
- [24] M.M. Rao and Z.D. Ren. *Theory of Orlicz spaces*, volume 146 of *Pure and Applied Mathematics*. Marcel Dekker, Inc., 1991.
- [25] R.T. Rockafellar. Integrals which are convex functionals. *Pacific J. Math.*, 24(3):525–539, 1968.
- [26] R.T. Rockafellar. *Conjugate Duality and Optimization*, volume 16 of *Regional Conferences Series in Applied Mathematics*. SIAM, Philadelphia, 1974.
- [27] R.T. Rockafellar and R. Wets. *Variational Analysis*, volume 317 of *Grundlehren der Mathematischen Wissenschaften*. Springer, 1998.

MODAL-X, UNIVERSITÉ PARIS 10 & BÂT. G, 200 AV. DE LA RÉPUBLIQUE. 92001 NANTERRE
CEDEX, FRANCE

CMAF, ÉCOLE POLYTECHNIQUE. 91128 PALAISEAU CEDEX, FRANCE
E-mail address: `christian.leonard@polytechnique.fr`