J. Math. Anal. Appl. 346 (2008) 183–204

Contents lists available at ScienceDirect

J. Math. Anal. Appl.

www.elsevier.com/locate/jmaa

Minimization of entropy functionals

Christian Léonard

Modal-X. Université Paris Ouest., Bât. G, 200 av. de la République, 92001 Nanterre, France

ARTICLE INFO

Article history:
Received 10 January 2008
Available online 26 April 2008
Submitted by A. Dontchev

Keywords:
Entropy
Convex optimization
Constraint qualification
Convex conjugate
Orlicz spaces

ABSTRACT

Entropy (i.e. convex integral) functionals and extensions of these functionals are minimized on convex sets. This paper is aimed at reducing as much as possible the assumptions on the constraint set. Primal attainment, dual equalities, dual attainment and characterizations of the minimizers are obtained with weak constraint qualifications. These results improve several aspects of the literature on the subject.

Contents

1. Introduction

Let $\mathcal{Z}$ be a space equipped with a positive reference measure R and denote $M_{\mathcal{Z}}$ the space of all signed measures on $\mathcal{Z}$. This paper is concerned with the minimization problem

$$\textit{minimize } I(Q) \quad \textit{subject to} \quad T_o Q \in C, \quad Q \in M_{\mathcal{Z}}, \tag{1.1}$$

where $T_o : M_{\mathcal{Z}} \to \mathcal{X}_o$ is a linear operator on $M_{\mathcal{Z}}$ which takes its values in a vector space $\mathcal{X}_o$, C is a convex subset of $\mathcal{X}_o$ and I is the *entropy* (i.e. convex integral) functional:

$$I(Q) = \begin{cases} \int_{\mathcal{Z}} \gamma_z^*(\frac{dQ}{dR}(z))\, R(dz) & \text{if } Q \prec R, \\ +\infty & \text{otherwise,} \end{cases} \quad Q \in M_{\mathcal{Z}}, \tag{1.2}$$

E-mail address: Christian.Leonard@u-paris10.fr.

0022-247X/$ – see front matter
doi:10.1016/j.jmaa.2008.04.048

where $Q \prec R$ means that Q is absolutely continuous with respect to R and γ^* is a $[0,\infty]$-valued measurable function on $\mathcal{Z} \times \mathbb{R}$ such that $\gamma^*(z,\cdot) := \gamma_z^*$ is convex and lower semicontinuous for all $z \in \mathcal{Z}$. Assume that for each z there exists a unique $m(z)$ which minimizes γ_z^* with

$$\gamma_z^*\big(m(z)\big) = 0, \quad \forall z \in \mathcal{Z}. \tag{1.3}$$

Then, I is $[0,\infty]$-valued, its unique minimizer is mR and $I(mR) = 0$. The solutions to (1.1) can be interpreted as I-projections of mR on $T_o^{-1}C$.

For some sets C such that (1.1) is not attained, it may still happen that an extended version of (1.1), which is stated at (1.11) below, is attained. This situation is also examined.

Presentation of the results

Our aim is to reduce as much as possible the restrictions on the constraint set C to obtain about (1.1) and its extension (1.11)

- an attainment criterion,
- dual equalities and
- a characterization of the minimizers.

Our main results are Theorems 4.2 and 3.2. Their proofs are based on abstract results which have been obtained by the author in [20]. They are exemplified at Section 7.

These results improve several aspects of the literature on the subject.

Clearly, for the problem (1.1) to be attained, $T_o^{-1}C$ must share a supporting hyperplane with some level set of I. This is the reason why it will be assumed to be closed with respect to some topology connected with the "geometry" of I. It will be the only restriction to be kept together with the interior specification (1.5) below.

Dual equalities and primal attainment are obtained under the weakest possible assumption:

$$C \cap T_o \operatorname{dom} I \neq \emptyset \tag{1.4}$$

where $T_o \operatorname{dom} I$ is the image by T_o of $\operatorname{dom} I := \{Q \in M_{\mathcal{Z}};\ I(Q) < \infty\}$. One obtains the characterization of the minimizers of (1.1) in the *interior case* which is specified by

$$C \cap \operatorname{icor}(T_o \operatorname{dom} I) \neq \emptyset \tag{1.5}$$

where $\operatorname{icor}(T_o \operatorname{dom} I)$ is the intrinsic core of $T_o \operatorname{dom} I$. The notion of intrinsic core does not rely on any topology; it gives the largest possible interior set. For comparison, a usual form of constraint qualification required for the representation of the minimizers of (1.1) is

$$\operatorname{int}(C) \cap T_o \operatorname{dom} I \neq \emptyset \tag{1.6}$$

where $\operatorname{int}(C)$ is the interior of C with respect to a topology which is not directly connected to I. In particular, $\operatorname{int}(C)$ must be nonempty; this is an important restriction. The constraint qualification (1.5) is weaker.

Similarly, a characterization of the solutions to the extended problem (1.11) is obtained together with dual equalities and a primal attainment criterion, under constraint qualifications analogous to (1.4) and (1.5) where I is replaced by $\bar{I}$.

Examples

A well-known entropy is the *relative entropy* with respect to the reference probability measure R which is defined on the set of all probability measure $P_{\mathcal{Z}}$ on $\mathcal{Z}$ by

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

It is close to the Boltzmann entropy (see Section 7.1) and corresponds to $\gamma_z^*(t) = t\log t - t + 1 + \iota_{\{t \geqslant 0\}}$ for all $z \in \mathcal{Z}$, where ι_A is the indicator of A, see (1.12). Inverting the roles of P and R one obtains the reverse relative entropy $I(R|P)$, $P \in P_{\mathcal{Z}}$. Popular entropies are

- the relative entropy: $\gamma^*(t) = t\log t - t + 1 + \iota_{\{t \geqslant 0\}}$,
- the reverse relative entropy: $\gamma^*(t) = t - \log t - 1 + \iota_{\{t > 0\}}$,
- the Fermi–Dirac entropy: $\gamma^*(t) = (1+t)\log(1+t) + (1-t)\log(1-t) + \iota_{\{-1 \leqslant t \leqslant 1\}}$,

- the L_p norm ($1 < p < \infty$): $\gamma^*(t) = |t|^p/p$ and
- the L_p entropy ($1 < p < \infty$): $\gamma^*(t) = t^p/p + \iota_{\{t \geqslant 0\}}$.

In the case of the relative and reverse relative entropies, the global minimizer is R which corresponds to $m = 1$ and one can interpret $I(P|R)$ and $I(R|P)$ as some type of distances between P and R. Note also that the positivity of the minimizers of (1.1) is guaranteed by $\operatorname{dom}\gamma^* \subset [0, \infty)$ where $\operatorname{dom}\gamma^*$ is the effective domain of γ^*. Consequently, the restriction that P is a probability measure is insured by the only unit mass constraint $P(\mathcal{Z}) = 1$.

The simplest constraint $T_o Q \in C$ is the moment constraint $\int_{\mathcal{Z}} \theta(z) Q(dz) \in C$ where θ is some numerical measurable function on $\mathcal{Z}$ and C is a real interval. Clearly $Q(\mathcal{Z}) = 1$ is a moment constraint. Considering K moment constraints simultaneously corresponds to the operator

$$T_o Q = \left(\int_{\mathcal{Z}} \theta_k(z) Q(dz) \right)_{1 \leqslant k \leqslant K} \in \mathbb{R}^K, \quad Q \in M_{\mathcal{Z}}, \tag{1.7}$$

where $\theta = (\theta_k)_{1 \leqslant k \leqslant K}$ is a $\mathbb{R}^K$-valued measurable function on $\mathcal{Z}$. As a typical instance of moment constraints, taking for $(\theta_k)_{1 \leqslant k \leqslant K}$ the first trigonometric polynomials gives a constraint on the first Fourier coefficients of Q. Infinitely many such constraints corresponds to the Fredholm integral operator

$$T_o Q = \left(\int_{\mathcal{Z}} \theta(r, z) Q(dz) \right)_{r \in \mathcal{R}} \in \mathbb{R}^{\mathcal{R}}, \quad Q \in M_{\mathcal{Z}}, \tag{1.8}$$

where θ is a real measurable function on $\mathcal{R} \times \mathcal{Z}$.

Another interesting example of constraint is the marginal constraint. Let $\mathcal{Z} = \mathcal{Z}_0 \times \mathcal{Z}_1$ be the product of the spaces $\mathcal{Z}_0$ and $\mathcal{Z}_1$. Let $P_0 \in P_{\mathcal{Z}_0}$ and $P_1 \in P_{\mathcal{Z}_1}$ denote the marginal measures on $\mathcal{Z}_0$ and $\mathcal{Z}_1$ of any probability measure $P \in P_{\mathcal{Z}_0 \times \mathcal{Z}_1}$. The prescribed marginal constraint corresponds to

$$T_o P = (P_0, P_1) \in P_{\mathcal{Z}_0} \times P_{\mathcal{Z}_1}, \quad P \in P_{\mathcal{Z}_0 \times \mathcal{Z}_1}. \tag{1.9}$$

This type of constraints occurs in tomography and image reconstruction where the solution of (1.1) is called a best entropy solution of the ill-posed inverse problem: "Find $P \in P_{\mathcal{Z}_0 \times \mathcal{Z}_1}$ such that P is absolutely continuous with respect to R, $P_0 = p_0$ and $P_1 = p_1$," p_0 and p_1 being prescribed marginals.

Taking $\mathcal{Z} = Z^{[0,1]}$ to be the set of all Z-valued paths $z = (z_t)_{0 \leqslant t \leqslant 1}$ (typically Z is $\mathbb{R}^d$ or a Riemannian manifold), a probability $P \in P_{\mathcal{Z}}$ is the law of a stochastic process. The standard stochastic calculus (Itô's formula) is kinetic. If one wishes to develop a dynamic approach to stochastic calculus, one is lead to least action principles which correspond to entropy minimization problems (1.1) on $P_{Z^{[0,1]}}$ with I the relative entropy under the marginal constraints

$$T_o P = (P_0, P_1) \in P_Z \times P_Z, \quad P \in P_{Z^{[0,1]}},$$

where P_0 and P_1 are the laws of the initial ($t = 0$) and final ($t = 1$) positions when P is the probability law of the whole path, see [9,14,28].

The extended minimization problem

A solution to (1.1) is in $A_{\mathcal{Z}}$: the space of all $Q \in M_{\mathcal{Z}}$ which are absolutely continuous with respect to R. One considers an extension $\bar{I}$ of the entropy I to a vector space $L_{\mathcal{Z}}$ which is the direct sum $L_{\mathcal{Z}} = A_{\mathcal{Z}} \oplus S_{\mathcal{Z}}$ of $A_{\mathcal{Z}}$ and a vector space $S_{\mathcal{Z}}$ of *singular linear forms* (acting on numerical functions) which may not be σ-additive. Any ℓ in $L_{\mathcal{Z}}$ is uniquely decomposed into $\ell = \ell^a + \ell^s$ with $\ell^a \in A_{\mathcal{Z}}$ and $\ell^s \in S_{\mathcal{Z}}$ and $\bar{I}$ has the following shape

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

where I^s is a positively homogeneous function on $S_{\mathcal{Z}}$. See (2.12) for the precise description of $\bar{I}$. For instance, the extended relative entropy is

$$\bar{I}(\ell|R) = I(\ell^a|R) + \sup\left\{ \langle \ell^s, u \rangle;\ u, \int_{\mathcal{Z}} e^u \, dR < \infty \right\}, \quad \ell \in L_{\mathcal{Z}}, \tag{1.10}$$

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

The extended problem is

$$\textit{minimize } \bar{I}(\ell) \quad \textit{subject to} \quad T_o \ell \in C, \quad \ell \in L_{\mathcal{Z}}. \tag{1.11}$$

Its precise statement is given at Section 2.3. In fact, $\bar{I}$ is chosen to be the largest convex lower semicontinuous extension of I to $L_{\mathcal{Z}}$ with respect to some weak topology. This guarantees tight relations between (1.1) and (1.11). In particular, one can expect that their values are equal for a large class of convex sets C.

Even if I is strictly convex, $\bar{I}$ is not strictly convex in general since I^s is positively homogeneous, so that (1.11) may admit several minimizers.

There are interesting situations where (1.1) is not attained in $A_{\mathcal{Z}}$ while (1.11) is attained in $L_{\mathcal{Z}}$. Examples are given at Section 7. This phenomenon is investigated in details by the author in [21] with probabilistic motivations.

Literature about entropy minimization

The maximum entropy method appear in many areas of applied mathematics and sciences such as statistical physics, information theory, mathematical statistics, large deviation theory, signal reconstruction and tomography. The literature is considerable: many papers are concerned with an engineering approach, working on the implementation of numerical procedures in specific situations. Entropy minimization is a popular method for solving ill-posed inverse problems, see [16,29] for instance.

Although entropy minimization has a long history, rigorous general results are rather recent. Let us cite, among others, the main contribution of Borwein and Lewis: [1–7] together with the paper [30] by Teboulle and Vajda. A Bayesian interpretation in the spirit of Jaynes [17] is developed by Dacunha-Castelle, Gamboa and Gassiat [13,15], see also [12]. In these papers, topological constraint qualifications of the type of (1.6) are required and results under the weaker constraint qualification (1.5) are obtained under the additional restriction that T_o has a finite dimensional rank (finitely many moment constraints). Such restrictions are removed here.

With a geometric point of view, Csiszár [10,11] provides a complete treatment of (1.1) with the relative entropy.

By means of a method different from the present one, the author has already studied in [23,24] entropy minimization problems under affine constraints (corresponding to C reduced to a single point) and more restrictive assumptions on γ^*.

The present article extends these results.

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

The intrinsic core of a subset A of a vector space is $\operatorname{icor} A=\{x\in A;\ \forall x'\in\operatorname{aff}A,\ \exists t>0,\ [x,x+t(x'-x)[\subset A\}$ where $\operatorname{aff}A$ is the affine space spanned by A. $\operatorname{icordom} f$ is the intrinsic core of the effective domain of f: $\operatorname{dom} f=\{x\in X;\ f(x)<\infty\}$. The indicator of a subset A of X is defined by

$$\iota_A(x)=\begin{cases}0 & \text{if } x\in A,\\ +\infty & \text{otherwise,}\end{cases}\qquad x\in X. \tag{1.12}$$

The support function of $A\subset X$ is $\iota_A^*(y)=\sup_{x\in A}\langle x,y\rangle$, $y\in Y$.

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 (1.2).

Outline of the paper

The minimization problems (1.1) and (1.11) are described in details at Section 2. Section 3 is devoted to the problem (1.1) and Section 4 to its extension (1.11). In Section 5, the main results of [20] about the extended saddle-point method are recalled. They are preliminary results for the proofs of Theorems 3.2 and 4.2 at Section 6. One presents important examples of entropies and constraints at Section 7.

2. The primal problems

The entropy minimization problems (1.1) and (1.11) are rigorously stated and renamed as (P_C) and $(\bar{\mathrm{P}}_C)$. Their correct mathematical statements necessitate the notion of Orlicz spaces.

The definitions of good and critical constraints are given and the main assumptions are collected at the end of this section.

2.1. Orlicz spaces

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 $\mathcal{L}_\rho$, $\mathcal{E}_\rho$, L_ρ and E_ρ).* The *Orlicz space* associated with ρ is defined by $\mathcal{L}_\rho(\mathcal{Z},R)=\{u:\mathcal{Z}\to\mathbb{R};\ \|u\|_\rho<+\infty\}$ where the Luxemburg norm $\|\cdot\|_\rho$ is defined by $\|u\|_\rho=\inf\{\beta>0;\ \int_{\mathcal{Z}}\rho(z,u(z)/\beta)R(dz)\leqslant 1\}$. Hence,

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

A subspace of interest is

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

Now, let us identify the R-a.e. equal functions. The corresponding spaces of equivalence classes are denoted $L_\rho(\mathcal{Z},R)$ and $E_\rho(\mathcal{Z},R)$.

Of course $E_\rho\subset L_\rho$. Note that if ρ does not 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 does not 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^*}.

Theorem 2.2 *(Representation of E'_ρ). Suppose that ρ is a finite Young function. Then, the dual space of E_ρ is isomorphic to L_{ρ^*}.*

Proof. For a proof of this result, see [18, Theorem 4.8]. □

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 $R(\bigcap_n A_n)=0$ and for all $n\geqslant 1$, $\langle\ell,u\mathbf{1}_{\mathcal{Z}\setminus A_n}\rangle=0$. Let us denote L^s_ρ the subspace of L'_ρ of all singular forms.

Theorem 2.3 *(Representation of L'_ρ). Let ρ be any Young function. The dual space of L_ρ is isomorphic to the direct sum $L'_\rho=(L_{\rho^*}\cdot R)\oplus L^s_\rho$. This implies that any $\ell\in L'_\rho$ is uniquely decomposed as*

$$\ell=\ell^a+\ell^s \tag{2.4}$$

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

Proof. When $L_\rho=L_\infty$ this result is the usual representation of L'_∞. When ρ is a finite function, this result is [19, Theorem 2.2]. The general result is proved in [25], with ρ not depending on z but the extension to a z-dependent ρ is obvious. □

In the decomposition (2.4), ℓ^a is called the *absolutely continuous* part of ℓ while ℓ^s is its *singular part*. The space L^s_ρ is the annihilator of E_ρ: $\ell\in L'_\rho$ is singular if and only if $\langle u,\ell\rangle=0$ for all $u\in E_\rho$.

Proposition 2.5. *Let us assume that ρ is finite. Then, $\ell\in L'_\rho$ is singular if and only if $\langle\ell,u\rangle=0$, for all u in E_ρ.*

Proof. This result is [19, Proposition 2.1]. □

The function ρ is said to satisfy the Δ_2-condition if

$$\text{there exist }\kappa>0,\ s_o\geqslant 0\quad\text{such that}\quad\forall s\geqslant s_o,\ z\in\mathcal{Z},\quad\rho_z(2s)\leqslant\kappa\rho_z(s). \tag{2.6}$$

If $s_o=0$, the Δ_2-condition is said to be global. When R is bounded, in order that $E_\rho=L_\rho$, it is enough that ρ satisfies the Δ_2-condition. When R is unbounded, this equality still holds if the Δ_2-condition is global. Consequently, if ρ satisfies the Δ_2-condition we have $L'_\rho=L_{\rho^*}\cdot R$ so that L^s_ρ reduces to the null vector space.

2.2. The minimization problem (P_C)

Let us state properly the basic problem (1.1).

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

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

where m satisfies (1.3), one sees that for R-a.e. z, λ_z is a nonnegative convex function and it vanishes at 0. Hence,

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

is a Young function. We shall use Orlicz spaces associated with $\lambda_\diamond$ and $\lambda_\diamond^*$.

We denote the space of R-absolutely continuous signed measures having a density in the Orlicz space $L_{\lambda_\diamond^*}$ by $L_{\lambda_\diamond^*}R$. The effective domain of I is included in $mR+L_{\lambda_\diamond^*}R$.

Constraint

In order to define the constraint, take $\mathcal{X}_o$ a vector space and a function $\theta:\mathcal{Z}\to\mathcal{X}_o$. One wants to give a meaning to the formal constraint $\int_{\mathcal{Z}}\theta\,dQ=x$ with $Q\in L_{\lambda_\diamond^*}R$ 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 $y\in\mathcal{Y}_o$,

$$T_o^*y(z):=\bigl\langle y,\theta(z)\bigr\rangle_{\mathcal{Y}_o,\mathcal{X}_o}, \quad z\in\mathcal{Z}. \tag{2.7}$$

Assuming that

$$T_o^*\mathcal{Y}_o\subset\mathcal{L}_{\lambda_\diamond}, \tag{2.8}$$

Hölder's inequality in Orlicz spaces allows to define the constraint operator $T_o\ell:=\int_{\mathcal{Z}}\theta\,d\ell$ for each $\ell\in L_{\lambda_\diamond^*}R$ by

$$\left\langle y,\int_{\mathcal{Z}}\theta\,d\ell\right\rangle_{\mathcal{Y}_o,\mathcal{X}_o}=\int_{\mathcal{Z}}\bigl\langle y,\theta(z)\bigr\rangle_{\mathcal{Y}_o,\mathcal{X}_o}\,\ell(dz), \quad \forall y\in\mathcal{Y}_o. \tag{2.9}$$

Minimization problem

Consider the minimization problem

$$\textit{minimize } I(Q) \quad \textit{subject to } \int_{\mathcal{Z}}\theta\,d(Q-mR)\in C_o, \quad Q\in mR+L_{\lambda_\diamond^*}R, \tag{P_{C_o}}$$

where C_o is a convex subset of $\mathcal{X}_o$. One sees with $\gamma_z^*(t)=\lambda_z^*(t-m(z))$ that $I_{\gamma^*}(Q)=I_{\lambda^*}(Q-mR)$. Therefore, the problem (P_{C_o}) is equivalent to

$$\textit{minimize } I_{\lambda^*}(\ell) \quad \textit{subject to } \int_{\mathcal{Z}}\theta\,d\ell\in C_o, \quad \ell\in L_{\lambda_\diamond^*}R, \tag{2.10}$$

with $\ell=Q-mR$. If the function m satisfies $m\in L_{\lambda_\diamond^*}$, one sees with (2.8) and Hölder's inequality in Orlicz spaces that the vector $x_o=\int_{\mathcal{Z}}\theta m\,dR\in\mathcal{X}_o$ is well defined in the weak sense. Therefore, (P_{C_o}) is

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

with $C=x_o+C_o$.

2.3. The extended minimization problem $(\bar{\mathrm{P}}_C)$

Let us state properly the extended problem (1.11). If the Young function $\lambda_\diamond$ does not satisfy the Δ_2-condition (2.6), for instance if it has an exponential growth at infinity as in (7.1) or even worse as in (7.4), the *small* Orlicz space $\mathcal{E}_{\lambda_\diamond}$ may be a proper subset of $\mathcal{L}_{\lambda_\diamond}$. Consequently, for some functions θ, the integrability property

$$T_o^*\mathcal{Y}_o\subset\mathcal{E}_{\lambda_\diamond} \tag{2.11}$$

or equivalently

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

may not be satisfied while the weaker property (2.8): $T_o^*\mathcal{Y}_o\subset\mathcal{L}_{\lambda_\diamond}$, or equivalently

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

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

If the constraint is critical, it may happen that (P_C) is not attained in $L_{\lambda_\diamond^*}R$. This is the reason why it is worth introducing its extension ($\bar{\mathrm{P}}_C$) which may admit minimizers and is defined by

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

where $L'_{\lambda_\diamond}$ is the topological dual space of $L_{\lambda_\diamond}$, $\bar{I}$ and $\langle\theta,\ell\rangle$ are defined below.

The dual space $L'_{\lambda_\diamond}$ admits the representation $L'_{\lambda_\diamond} \simeq L_{\lambda_\diamond^*}R \oplus L^s_{\lambda_\diamond}$. This means that any $\ell \in L'_{\lambda_\diamond}$ is uniquely decomposed as $\ell = \ell^a + \ell^s$ where $\ell^a \in L_{\lambda_\diamond^*}R$ and $\ell^s \in L^s_{\lambda_\diamond}$ are respectively the *absolutely continuous part* and the *singular part* of ℓ, see Theorem 2.3. The extension $\bar{I}$ has the following form

$$\bar{I}(\ell) = I\big(\ell^a\big) + \iota^*_{\mathrm{dom}\, I_\gamma}\big(\ell^s\big), \quad \ell \in L'_{\lambda_\diamond}. \tag{2.12}$$

It will be shown that $\bar{I}$ is the greatest convex $\sigma(L'_{\lambda_\diamond}, L_{\lambda_\diamond})$-lower semicontinuous extension of I to $L'_{\lambda_\diamond} \supset L_{\lambda_\diamond^*}$. In a similar way to (2.9), the hypothesis ($A_\theta^\exists$) allows to define

$$T\ell := \langle\theta,\ell\rangle$$

for all $\ell \in L'_{\lambda_\diamond}$ by

$$\big\langle y, \langle\theta,\ell\rangle\big\rangle_{\mathcal{Y}_o,\mathcal{X}_o} = \big\langle\langle y,\theta\rangle,\ell\big\rangle_{L_{\lambda_\diamond},L'_{\lambda_\diamond}}, \quad \forall y \in \mathcal{Y}_o.$$

Important examples of entropies with $\lambda_\diamond$ not satisfying the Δ_2-condition are the usual (Boltzmann) entropy and its variants, see Section 7.1 and (7.1) in particular.

When $\lambda_\diamond$ satisfies the Δ_2-condition (2.6), ($\bar{\mathrm{P}}_C$) is (P_C).

2.4. Hypotheses

Let us collect the hypotheses on R, γ^* and θ.

Hypotheses (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^*}) *Hypotheses 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 its (unique) minimum at $m(z)$ with $\gamma^*(z,m(z)) = 0$.

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

(A_θ) *Hypotheses 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 $y = 0$;

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

Remarks 2.13. Some technical remarks about the hypotheses.

(a) Since γ_z^* is a convex function on $\mathbb{R}$, it is continuous on the interior of its domain. Under our hypotheses, γ^* is (jointly) measurable, and so are γ and m. Hence, λ is also measurable.

(b) As γ_z^* is strictly convex, γ_z is differentiable on the interior of its effective domain.

(c) Hypothesis ($\mathrm{A}^2_{\gamma^*}$) is $m \in L_{\lambda_\diamond^*}$. It allows to consider problem (P_C) rather than (P_{C_o}). If this hypothesis is not satisfied, our results still hold for (P_{C_o}), but their statement is a little heavier, see Remark 4.6(d) below.

(d) Since $\mathcal{X}_o$ and $\mathcal{Y}_o$ are in separating duality, (A^2_θ) states that the vector space spanned by the range of θ "is essentially" $\mathcal{X}_o$. This is not an effective restriction.

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

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

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

The space $\mathcal{X}$. The topological dual space of $\mathcal{Y}$ is $\mathcal{X} = \mathcal{Y}' \subset \mathcal{X}_o$. It will be shown at (6.7) that under our assumptions $TL'_{\lambda_\diamond} \subset \mathcal{X}$. Hence, $\mathcal{X}$ is identified with $L'_{\lambda_\diamond}/\ker T$ and its norm is given by $|x|^*_\Lambda = \inf\{\|\ell\|^*_{\lambda_\diamond};\ \ell \in L'_{\lambda_\diamond}\colon\ T\ell = x\}$.

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

Note that under the assumption $(\mathrm{A}_\theta^\forall)$, $T^*\mathcal{Y} \subset E_{\lambda_\diamond}$ and $T(L_{\lambda_\diamond^*}R) \subset \mathcal{X}$.

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

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

where $\Gamma(y) = I_\gamma(\langle y, \theta\rangle)$, $y \in \mathcal{Y}_o$. Let us denote

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

its convex conjugate. It will be shown at (6.8) that $\operatorname{dom}\Gamma^* \subset \mathcal{X}$.

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

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

3. Solving (P_C)

In this section, the general hypotheses (A) are imposed and we study (P_C) under the additional good constraint hypothesis $(\mathrm{A}_\theta^\forall)$ which imposes that $T^*\mathcal{Y} \subset E_{\lambda_\diamond}$.

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

$$\begin{cases} \lambda_+(z, s) = \lambda\big(z, |s|\big), \\ \lambda_-(z, s) = \lambda\big(z, -|s|\big), \end{cases} \tag{3.1}$$

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

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

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

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

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

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

Theorem 3.2. *Suppose that*

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

$$T_o^{-1}C \cap L_{\lambda_\diamond^*}R = \bigcap_{y\in Y}\left\{fR \in L_{\lambda_\diamond^*}R;\ \int_{\mathcal{Z}} \langle y, \theta\rangle f\,dR \geqslant a_y\right\} \tag{3.3}$$

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

Then:

(a) *The dual equality for* (P_C) *is*

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

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

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

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

$$\begin{cases} \text{(a)} \quad \hat{x} \in C \cap \operatorname{dom} \Gamma^*, \\ \text{(b)} \quad \langle \tilde{\omega}, \hat{x} \rangle_{\mathcal{X}^*, \mathcal{X}} \leqslant \langle \tilde{\omega}, x \rangle_{\mathcal{X}^*, \mathcal{X}}, \quad \forall x \in C \cap \operatorname{dom} \Gamma^*, \\ \text{(c)} \quad \widehat{Q}(dz) = \gamma_z'\big(\langle \tilde{\omega}, \theta(z) \rangle\big) R(dz). \end{cases} \tag{3.4}$$

Furthermore, $\widehat{Q} \in L_{\lambda_\diamond^*} R$ *and* $\tilde{\omega} \in \widetilde{\mathcal{Y}}$ *satisfy* (3.4) *if and only if* $\widehat{Q}$ *solves* (P_C) *and* $\tilde{\omega}$ *solves* $(\widetilde{\mathrm{D}}_C)$.

(d) *Of course,* (3.4(c)) *implies* $\hat{x} = \int_{\mathcal{Z}} \theta \gamma'(\langle \tilde{\omega}, \theta \rangle) \, dR$ *in the weak sense. Moreover,*
1. $\hat{x}$ *minimizes* Γ^* *on* C,
2. $I(\widehat{Q}) = \Gamma^*(\hat{x}) = \int_{\mathcal{Z}} \gamma^* \circ \gamma'(\langle \tilde{\omega}, \theta \rangle) \, dR < \infty$ *and*
3. $I(\widehat{Q}) + \int_{\mathcal{Z}} \gamma(\langle \tilde{\omega}, \theta \rangle) \, dR = \int_{\mathcal{Z}} \langle \tilde{\omega}, \theta \rangle \, d\widehat{Q}$.

Remarks 3.5.

(a) Suppose that γ^* does not depend on z (to simplify) and $\liminf_{|t| \to +\infty} \gamma^*(t)/|t| < +\infty$, then λ is not a finite function, $E_{\lambda_\diamond} = \{0\}$ and $(\mathrm{A}_\theta^\forall)$ implies that $\theta = 0$.

(b) Removing the hypothesis $(\mathrm{A}^2_{\gamma^*})$: $m \in L_{\lambda_\diamond^*}$, one can still consider the minimization problem (P_{C_o}) instead of (P_C). The transcription of Theorem 3.2 is as follows. Replace respectively (P_C), C, Γ^*, $\hat{x}$ and γ by (P_{C_o}), C_o, Λ^*, $\tilde{x}$ and λ where $\tilde{x} = \int_{\mathcal{Z}} \theta \, d(\widehat{Q} - mR)$ is well defined.
The statement (b) must be replaced by the following one: *If $C_o \cap \operatorname{dom} \Lambda^* \neq \emptyset$, then (P_{C_o}) admits a unique solution $\widehat{Q}$ in $mR + L_{\lambda_\diamond^*} R$ and any minimizing sequence $(Q_n)_{n \geqslant 1}$ is such that $(Q_n - mR)_{n \geqslant 1}$ converges in $L_{\lambda_\diamond^*} R$ to $\widehat{Q} - mR$ with respect to the topology $\sigma(L_{\lambda_\diamond^*} R, L_{\lambda_\diamond})$.*

(c) For comparison with (3.3), note that the general shape of $T_o^{-1} C \cap L_{\lambda_\diamond^*} R$ when C is only supposed to be convex is $\bigcap_{(y,a) \in A} \{\ell \in L_{\lambda_\diamond^*} R;\ \langle \langle y, \theta \rangle, \ell \rangle > a\}$ with $A \subset \mathcal{Y} \times \mathbb{R}$.

4. Solving $(\bar{\mathrm{P}}_C)$

The general hypotheses (A) are imposed and we study $(\bar{\mathrm{P}}_C)$. Again, one needs to introduce several cones to state the dual problems.

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

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

A decomposition in K''_λ. Let ρ be any Young function. By Theorem 2.3, we have $L''_\rho = [L_\rho \oplus L^s_{\rho^*}] \oplus L^{s\prime}_\rho$. For any $\zeta \in L''_\rho = (L_{\rho^*} R \oplus L^s_\rho)'$, let us denote the restrictions $\zeta_1 = \zeta_{|L_{\rho^*} R}$ and $\zeta_2 = \zeta_{|L^s_\rho}$. Since, $(L_{\rho^*} R)' \simeq L_\rho \oplus L^s_{\rho^*}$, one sees that any $\zeta \in L''_\rho$ is uniquely decomposed into

$$\zeta = \zeta_1^a + \zeta_1^s + \zeta_2 \tag{4.1}$$

with $\zeta_1 = \zeta_1^a + \zeta_1^s \in L'_{\rho^*}$, $\zeta_1^a \in L_\rho$, $\zeta_1^s \in L^s_{\rho^*}$ and $\zeta_2 \in L^{s\prime}_\rho$. Translating this decomposition onto K''_λ leads to $K''_\lambda = [K_\lambda \oplus K^s_{\lambda^*}] \oplus K^{s\prime}_\lambda$ where one defines $K^s_{\lambda^*} = \{\zeta \in (L_{\lambda_\diamond^*} R)^b;\ \zeta_{\pm | L_{\lambda_\pm^*} R} \in L^s_{\lambda_\pm^*}\}$ and $K^{s\prime}_\lambda = \{\zeta \in L^{sb}_{\lambda_\diamond};\ \zeta_{\pm | L^s_{\lambda_\pm}} \in L^{s\prime}_{\lambda_\pm}\}$. Note that $L_{\lambda_\pm^*} R \subset L_{\lambda_\diamond^*} R$ and $L^s_{\lambda_\pm} \subset L^s_{\lambda_\diamond}$. With these cones in hand, the decomposition (4.1) holds for any $\zeta \in K''_\lambda$ with

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

The set $\overline{\mathcal{Y}}$. The $\sigma(K''_\lambda, L'_\pm)$-closure $\overline{A}$ of a set $A \subset K''_\lambda$ is defined as follows: $\zeta \in K''_\lambda$ is in $\overline{A}$ if $\zeta_\pm$ is in the $\sigma(L''_{\lambda_\pm}, L'_{\lambda_\pm})$-closure of $A_\pm = \{\zeta_\pm;\ \zeta \in A\}$. Clearly, $\overline{A}_\pm = \{\zeta_\pm;\ \zeta \in \overline{A}\}$. Let $\overline{T_o^* \mathcal{Y}_o}$ denote the $\sigma(K''_\lambda, L'_\pm)$-closure of $T_o^* \mathcal{Y}_o$ in K''_λ.

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

The set $\overline{\mathcal{Y}} \subset \mathcal{X}^*$ is the extension of $\mathcal{Y}_o$ which is isomorphic to $\overline{T_o^* \mathcal{Y}_o} \cap \{\zeta \in K_\lambda'';\ \zeta_1^s = 0,\ \zeta_2 \in D\}$ in the sense that

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

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

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

Theorem 4.2. *Suppose that*

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

$$T_o^{-1} C \cap L_{\lambda_\diamond}' = \bigcap_{y \in Y} \{\ell \in L_{\lambda_\diamond}';\ \langle \langle y, \theta\rangle, \ell\rangle \geqslant a_y\} \tag{4.3}$$

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

Then:

(a) *The dual equality for $(\bar{\mathrm{P}}_C)$ is*

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

(b) *If $C \cap \operatorname{dom} \Gamma^* \neq \emptyset$ or equivalently $C \cap T_o \operatorname{dom} \bar{I} \neq \emptyset$, then $(\bar{\mathrm{P}}_C)$ admits solutions in $L_{\lambda_\diamond}'$, any minimizing sequence admits $\sigma(L_{\lambda_\diamond}', L_{\lambda_\diamond})$-cluster points and every such point is a solution to $(\bar{\mathrm{P}}_C)$.*

Suppose that in addition we have

$$C \cap \operatorname{icordom} \Gamma^* \neq \emptyset \tag{4.4}$$

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

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

$$\begin{cases} \text{(a)} & \hat{x} \in C \cap \operatorname{dom} \Gamma^*, \\ \text{(b)} & \langle \bar{\omega}, \hat{x}\rangle_{\mathcal{X}^*, \mathcal{X}} \leqslant \langle \bar{\omega}, x\rangle_{\mathcal{X}^*, \mathcal{X}}, \quad \forall x \in C \cap \operatorname{dom} \Gamma^*, \\ \text{(c)} & \hat{\ell} \in \gamma_z'\big([T^*\bar{\omega}]_1^a\big) R + D^\perp\big([T^*\bar{\omega}]_2\big) \end{cases} \tag{4.5}$$

where

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

is the outer normal cone of D at η.

$T^*\bar{\omega}$ is in the $\sigma(K_\lambda'', L_\pm')$-closure of $T^*(\operatorname{dom} \Lambda)$ and there exists some $\tilde{\omega} \in \mathcal{X}_o^*$ such that

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

is a measurable function in the strong closure of $T^*(\operatorname{dom} \Lambda)$ in K_λ: the set of all $u \in K_\lambda$ such that $u_\pm$ is in the strong closure of $T^*(\operatorname{dom} \Lambda)_\pm$ in $L_{\lambda_\pm}$.

Furthermore, $\hat{\ell} \in L_{\lambda_\diamond}'$ and $\bar{\omega} \in \overline{\mathcal{Y}}$ satisfy (4.5) *if and only if $\hat{\ell}$ solves $(\bar{\mathrm{P}}_C)$ and $\bar{\omega}$ solves $(\bar{\mathrm{D}}_C)$.*

(d) *Of course,* (4.5(c)) *implies $\hat{x} = \int_{\mathcal{Z}} \theta \gamma'(\langle \tilde{\omega}, \theta\rangle)\, dR + \langle \theta, \hat{\ell}^s\rangle$. Moreover,*
1. *$\hat{x}$ minimizes Γ^* on C,*
2. *$\bar{I}(\hat{\ell}) = \Gamma^*(\hat{x}) = \int_{\mathcal{Z}} \gamma^* \circ \gamma'(\langle \tilde{\omega}, \theta\rangle)\, dR + \sup\{\langle u, \hat{\ell}^s\rangle;\ u \in \operatorname{dom} I_\gamma\} < \infty$ and*
3. *$\bar{I}(\hat{\ell}) + \int_{\mathcal{Z}} \gamma(\langle \tilde{\omega}, \theta\rangle)\, dR = \int_{\mathcal{Z}} \langle \tilde{\omega}, \theta\rangle\, d\hat{\ell}^a + \langle [T^*\bar{\omega}]_2, \hat{\ell}^s\rangle_{K_\lambda^{s\prime}, K_\lambda^s}$.*

Remarks 4.6. General remarks about Theorem 4.2.

(a) The hypothesis (2) is equivalent to C is $\sigma(\mathcal{X}, \mathcal{Y})$-closed convex.

(b) The dual equality with $C = \{x\}$ gives for all $x \in \mathcal{X}_o$

$$\Gamma^*(x) = \inf\{\bar{I}(\ell);\ \ell \in L'_{\lambda_\diamond},\ \langle\theta, \ell\rangle = x\}.$$

(c) Note that $\bar{\omega}$ does not necessarily belong to $\mathcal{Y}_o$. Therefore, the Young equality $\langle\bar{\omega}, \hat{x}\rangle = \Gamma^*(\hat{x}) + \Gamma(\bar{\omega})$ is meaningless. Nevertheless, there exists a natural extension $\overline{\Gamma}$ of Γ such that $\langle\hat{x}, \bar{\omega}\rangle = \Gamma^*(\hat{x}) + \overline{\Gamma}(\bar{\omega})$ holds, see (5.6). This gives the statement (d)3.

(d) As in Remark 3.5(b), removing the hypothesis $(\mathrm{A}^2_{\gamma^*})$: $m \in L_{\lambda^*_\diamond}$, one can still consider the minimization problem

$$\textit{minimize } \bar{I}(\ell) \quad \textit{subject to} \quad \langle\theta, \ell - mR\rangle \in C_o, \quad \ell \in mR + L'_{\lambda_\diamond}, \tag{$\bar{\mathrm{P}}_{C_o}$}$$

instead of $(\bar{\mathrm{P}}_C)$. The transcription of Theorem 4.2 is as follows. Denote

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

and replace respectively $(\bar{\mathrm{P}}_C)$, C, Γ^*, $\hat{x}$ and γ by $(\bar{\mathrm{P}}_{C_o})$, C_o, Λ^*, $\tilde{x}$ and λ where $\tilde{x} = \langle\theta, \hat{\ell} - mR\rangle$ is well defined.
The statement (b) must be replaced by the following one: *If $C_o \cap \operatorname{dom}\Lambda^* \neq \emptyset$, then $(\bar{\mathrm{P}}_{C_o})$ admits solutions in $mR + L'_{\lambda_\diamond}$, any minimizing sequence $(\ell_n)_{n\geqslant 1}$ is such that $(\ell_n - mR)_{n\geqslant 1}$ admits cluster points $\hat{\ell} - mR$ in $L'_{\lambda^*_\diamond}$ with respect to the topology $\sigma(L'_{\lambda_\diamond}, L_{\lambda_\diamond})$ and $\hat{\ell}$ is a solution of $(\bar{\mathrm{P}}_{C_o})$.*

Example 4.7. Now, we give a simple example to illustrate Part (c) of Theorem 4.2. Consider three constraint functions θ_1, θ_2 and θ_3 and the problem

$$\textit{minimize } \bar{I}(\ell) \quad \textit{subject to} \quad (\langle\theta_1, \ell\rangle, \langle\theta_2, \ell\rangle, \langle\theta_3, \ell\rangle) \in C, \quad \ell \in L'_{\lambda_\diamond},$$

where C is a closed convex subset of $\mathbb{R}^3$ which satisfies (4.4). Let γ^* be such that $\lambda_\diamond$ is not Δ_2-regular, so that $E_{\lambda_\diamond} \subsetneq L_{\lambda_\diamond}$ and suppose that $\theta_1, \theta_2 \in L_{\lambda_\diamond} \setminus E_{\lambda_\diamond}$ while $\theta_3 \in E_{\lambda_\diamond}$.

We have $T_o^* y(z) = \sum_{1\leqslant k\leqslant 3} y_k\theta_k(z)$, $y = (y_1, y_2, y_3) \in \mathbb{R}^3$, $z \in \mathcal{Z}$, and the closure operations leading to $\overline{\mathcal{Y}}$ are trivial because of the finite dimension. Hence, in (4.5) we have $\bar{\omega} \in \mathbb{R}^3$ with $[T^*\bar{\omega}]_1^a = \bar{\omega}_1\theta_1 + \bar{\omega}_2\theta_2 + \bar{\omega}_3\theta_3$ and $[T^*\bar{\omega}]_2 = \bar{\omega}_1\theta_1 + \bar{\omega}_2\theta_2$.

The singular component $\hat{\ell}^s$ is in $D^\perp(\bar{\omega}_1\theta_1 + \bar{\omega}_2\theta_2)$. For $\hat{\ell}^s$ to be nonzero, it is necessary that $\bar{\omega}_1\theta_1 + \bar{\omega}_2\theta_2$ is on the boundary of $\operatorname{dom} I_\lambda$, that is $I_\lambda(t[\bar{\omega}_1\theta_1 + \bar{\omega}_2\theta_2]) < \infty$ for all $0 \leqslant t < 1$ and $I_\lambda(t[\bar{\omega}_1\theta_1 + \bar{\omega}_2\theta_2]) = \infty$ for all $t > 1$. As $\hat{\ell}^s$ is singular, for each $\epsilon > 0$ there exists a measurable set S_ϵ such that $R(S_\epsilon) < \epsilon$ and $\langle\hat{\ell}^s, u\mathbf{1}_{S^c_\epsilon}\rangle = 0$ for all $u \in L_{\lambda_\diamond}$. Since $\hat{\ell}^s \in D^\perp(\bar{\omega}_1\theta_1 + \bar{\omega}_2\theta_2)$, it is also necessary that

$$I_\lambda(t\mathbf{1}_{S_\epsilon}[\bar{\omega}_1\theta_1 + \bar{\omega}_2\theta_2]) = \infty, \quad t > 1. \tag{4.8}$$

This gives an information on the "support" of $\hat{\ell}^s$.

In the special case of the reverse relative entropy, $L_{\lambda_\diamond} = L_\infty$, $E_{\lambda_\diamond} = \{0\}$, $\operatorname{dom}\lambda_+ = (-1, +1)$ and λ_- is Δ_2-regular, see (7.4). Because of this special form, the blowup of $t \mapsto I_\lambda(t[\bar{\omega}_1\theta_1 + \bar{\omega}_2\theta_2])$ is easy to describe: We have $\bar{\omega}_1\theta_1 + \bar{\omega}_2\theta_2 \leqslant 1$ R-a.e. and for all $\epsilon > 0$, S_ϵ must meet $\{\bar{\omega}_1\theta_1 + \bar{\omega}_2\theta_2 = 1\}$ for $\hat{\ell}^s$ to be nonzero. With some abuse, one might say that the "support" of $\hat{\ell}^s$ is an R-negligible subset of $\{\bar{\omega}_1\theta_1 + \bar{\omega}_2\theta_2 = 1\}$.

At examples (b) and (c) at Section 7.4 which involve the relative entropy, a description of $\hat{\ell}^s$ is also given.

5. Preliminary results

The aim of this section is to recall for the convenience of the reader some results of [20,22,24].

5.1. The extended saddle-point method

Denoting the minimizer $\widehat{Q}$ of (P_C), the geometric picture is that some level set of I is tangent at $\widehat{Q}$ to the constraint set $T_o^{-1}C$. Since these sets are convex, they are separated by some affine hyperplane and the analytic description of this separation yields the characterization of $\widehat{Q}$. Of course Hahn–Banach theorem is the key. Standard approaches require C to be open with respect to some given topology in order to be allowed to apply it. In the present paper, one chooses to use a topological structure which is designed for the level sets of I to "look like" open sets, so that Hahn–Banach theorem can be applied without assuming to much on C.

This strategy is implemented in [20] in an abstract setting suitable for several applications. It is a refinement of the standard saddle-point method [27] where convex conjugates play an important role.

5.2. The failure of maximum entropy reconstruction

The main problem one has to overcome when working with infinite dimensional constraints is that the dual attainment is not the rule. Borwein [1] calls this phenomenon the failure of maximum entropy reconstruction and obtains representation of the minimizers in terms of approximating sequences. An aim of [20] was to provide tools to obtain general representations of the minimizers without approximation. This is done in the present paper where integral representations of the minimizers are derived.

5.3. Convex minimization problems under weak constraint qualifications

The main results of [20] are presented.

Basic diagram

Let $\mathcal{U}_o$ be a vector space, $\mathcal{L}_o = \mathcal{U}_o^*$ its algebraic dual space, Φ a $(-\infty, +\infty]$-valued convex function on $\mathcal{U}_o$ and Φ^* its convex conjugate for the duality $\langle \mathcal{U}_o, \mathcal{L}_o \rangle$:

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

Let $\mathcal{Y}_o$ be another vector space, $\mathcal{X}_o = \mathcal{Y}_o^*$ its algebraic dual space and $T_o : \mathcal{L}_o \to \mathcal{X}_o$ a linear operator. We consider the convex minimization problem

$$\textit{minimize } \Phi^*(\ell) \quad \textit{subject to} \quad T_o \ell \in C, \quad \ell \in \mathcal{L}_o, \tag{$\mathcal{P}_o$}$$

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

This will be used later with $\Phi = I_\lambda$ on the Orlicz space $\mathcal{U}_o = \mathcal{E}_{\lambda_\diamond}(\mathcal{Z}, R)$ or $\mathcal{U}_o = \mathcal{L}_{\lambda_\diamond}(\mathcal{Z}, R)$.

It is useful to define the constraint operator T_o by means of its adjoint $T_o^* : \mathcal{Y}_o \to \mathcal{L}_o^*$ for each $\ell \in \mathcal{L}_o$, by $\langle T_o^* y, \ell \rangle_{\mathcal{L}_o^*, \mathcal{L}_o} = \langle y, T_o \ell \rangle_{\mathcal{Y}_o, \mathcal{X}_o}$, $\forall y \in \mathcal{Y}_o$.

Hypotheses

Let us give the list of the main hypotheses.

(H_Φ) 1. $\Phi : \mathcal{U}_o \to [0, +\infty]$ is $\sigma(\mathcal{U}_o, \mathcal{L}_o)$-lower semicontinuous, convex and $\Phi(0) = 0$;
2. $\forall u \in \mathcal{U}_o$, $\exists \alpha > 0$, $\Phi(\alpha u) < \infty$;
3. $\forall u \in \mathcal{U}_o$, $u \neq 0$, $\exists t \in \mathbb{R}$, $\Phi(tu) > 0$.

(H_T) 1. $T_o^*(\mathcal{Y}_o) \subset \mathcal{U}_o$;
2. $\ker T_o^* = \{0\}$.

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

The definitions of the vector spaces $\mathcal{X}$ and $\mathcal{Y}$ which appear in the last hypothesis are stated below. For the moment, let us only say that if C is convex and $\sigma(\mathcal{X}_o, \mathcal{Y}_o)$-closed, then ($\mathrm{H}_C$) holds.

Several primal and dual problems

These variants are expressed below in terms of new spaces and functions. Let us first introduce them.

- The norms $|\cdot|_\Phi$ and $|\cdot|_\Lambda$. Let $\Phi_\pm(u) = \max(\Phi(u), \Phi(-u))$. By ($\mathrm{H}_{\Phi}1$) and ($\mathrm{H}_{\Phi}2$), $\{u \in \mathcal{U}_o;\ \Phi_\pm(u) \leqslant 1\}$ is a convex absorbing balanced set. Hence its gauge functional which is defined for all $u \in \mathcal{U}_o$ by $|u|_\Phi := \inf\{\alpha > 0;\ \Phi_\pm(u/\alpha)) \leqslant 1\}$ is a seminorm. Thanks to hypothesis ($\mathrm{H}_{\Phi}3$), it is a norm.
  Taking ($\mathrm{H}_{T}1$) into account, one can define
  $$\Lambda_o(y) := \Phi(T_o^* y), \quad y \in \mathcal{Y}_o. \tag{5.1}$$
  Let $\Lambda_\pm(y) = \max(\Lambda_o(y), \Lambda_o(-y))$. The gauge functional on $\mathcal{Y}_o$ of the set $\{y \in \mathcal{Y}_o;\ \Lambda_\pm(y) \leqslant 1\}$ is $|y|_\Lambda := \inf\{\alpha > 0;\ \Lambda_\pm(y/\alpha) \leqslant 1\}$, $y \in \mathcal{Y}_o$. Thanks to (H_Φ) and (H_T), it is a norm and
  $$|y|_\Lambda = |T_o^* y|_\Phi, \quad y \in \mathcal{Y}_o.$$
- The spaces. Let

  $\mathcal{U}$ be the $|\cdot|_\Phi$-completion of $\mathcal{U}_o$ and let

  $\mathcal{L} := (\mathcal{U}_o, |\cdot|_\Phi)'$ be the topological dual space of $(\mathcal{U}_o, |\cdot|_\Phi)$.

Of course, we have $(\mathcal{U}, |\cdot|_\Phi)' \cong \mathcal{L} \subset \mathcal{L}_o$ where any ℓ in $\mathcal{U}'$ is identified with its restriction to $\mathcal{U}_o$. Similarly, we introduce

$$\mathcal{Y} \text{ the } |\cdot|_\Lambda\text{-completion of } \mathcal{Y}_o \text{ and}$$
$$\mathcal{X} := \big(\mathcal{Y}_o, |\cdot|_\Lambda\big)' \text{ the topological dual space of } \big(\mathcal{Y}_o, |\cdot|_\Lambda\big).$$

We have $(\mathcal{Y}, |\cdot|_\Lambda)' \cong \mathcal{X} \subset \mathcal{X}_o$ where any x in $\mathcal{Y}'$ is identified with its restriction to $\mathcal{Y}_o$.
We also have to consider the *algebraic* dual spaces $\mathcal{L}^*$ and $\mathcal{X}^*$ of $\mathcal{L}$ and $\mathcal{X}$.

– The operators T and T^*. Let us denote T the restriction of T_o to $\mathcal{L} \subset \mathcal{L}_o$. One can show that under $(\mathrm{H}_{\Phi\&T})$,

$$T_o\mathcal{L} \subset \mathcal{X}. \tag{5.2}$$

Hence $T : \mathcal{L} \to \mathcal{X}$. Let us define its adjoint $T^* : \mathcal{X}^* \to \mathcal{L}^*$ for all $\omega \in \mathcal{X}^*$ by $\langle \ell, T^*\omega\rangle_{\mathcal{L},\mathcal{L}^*} = \langle T\ell, \omega\rangle_{\mathcal{X},\mathcal{X}^*}$, $\forall \ell \in \mathcal{L}$. We have the inclusions $\mathcal{Y}_o \subset \mathcal{Y} \subset \mathcal{X}^*$. The adjoint operator T_o^* is the restriction of T^* to $\mathcal{Y}_o$.

– The functionals. They are

$$\begin{aligned}
&\bar\Phi(\zeta) := \sup\nolimits_{\ell\in\mathcal{L}}\big\{\langle\zeta,\ell\rangle - \Phi^*(\ell)\big\}, && \zeta \in \mathcal{L}^*,\\
&\Lambda(y) := \bar\Phi\big(T^*y\big), && y \in \mathcal{Y},\\
&\overline{\Lambda}(\omega) := \bar\Phi\big(T^*\omega\big), && \omega \in \mathcal{X}^*,\\
&\Lambda_o^*(x) := \sup\nolimits_{y\in\mathcal{Y}_o}\big\{\langle y,x\rangle - \Lambda_o(y)\big\}, && x \in \mathcal{X}_o,\\
&\Lambda^*(x) := \sup\nolimits_{y\in\mathcal{Y}}\big\{\langle y,x\rangle - \Lambda(y)\big\}, && x \in \mathcal{X}.
\end{aligned}$$

– The optimization problems. They are

$$\begin{aligned}
&\textit{minimize } \Phi^*(\ell) && \textit{subject to } \ T_o\ell \in C, \quad \ell \in \mathcal{L}_o, && (\mathcal{P}_o)\\
&\textit{minimize } \Phi^*(\ell) && \textit{subject to } \ T\ell \in C, \quad \ell \in \mathcal{L}, && (\mathcal{P})\\
&\textit{maximize } \inf\nolimits_{x\in C\cap\mathcal{X}}\langle y,x\rangle - \Lambda(y), && \quad y \in \mathcal{Y}, && (\mathcal{D})\\
&\textit{maximize } \inf\nolimits_{x\in C\cap\mathcal{X}}\langle x,\omega\rangle - \overline{\Lambda}(\omega), && \quad \omega \in \mathcal{X}^*. && (\overline{\mathcal{D}})
\end{aligned}$$

Statement of the results

It is assumed that (H_Φ), (H_T) and (H_C) hold.

Theorem 5.3 *(Primal attainment and dual equality).*

(a) *The problems* $(\mathcal{P}_o)$ *and* $(\mathcal{P})$ *are equivalent: they have the same solutions and* $\inf(\mathcal{P}_o) = \inf(\mathcal{P}) \in [0,\infty]$. *In particular,* $\operatorname{dom}\Phi^* \subset \mathcal{L}$ *and* $\operatorname{dom}\Lambda^* \subset \mathcal{X}$.
(b) *We have the dual equalities*

$$\inf(\mathcal{P}_o) = \inf(\mathcal{P}) = \sup(\mathcal{D}) = \sup(\overline{\mathcal{D}}) = \inf_{x\in C}\Lambda_o^*(x) = \inf_{x\in C\cap\mathcal{X}}\Lambda^*(x) \in [0,\infty].$$

(c) *If in addition* $\{\ell \in \mathcal{L}_o;\ T_o\ell \in C\} \cap \operatorname{dom}\Phi^* \neq \emptyset$, *then* $(\mathcal{P}_o)$ *is attained in* $\mathcal{L}$. *Moreover, any minimizing sequence for* $(\mathcal{P}_o)$ *has* $\sigma(\mathcal{L},\mathcal{U})$*-cluster points and every such cluster point solves* $(\mathcal{P}_o)$.

Theorem 5.4 *(Dual attainment and representation. Interior convex constraint). Assume that* $C \cap \operatorname{icor}(T_o \operatorname{dom}\Phi^*) \neq \emptyset$. *Then, the primal problem* $(\mathcal{P}_o)$ *is attained in* $\mathcal{L}$ *and the extended dual problem* $(\overline{\mathcal{D}})$ *is attained in* $\mathcal{X}^*$. *Any solution* $\hat\ell \in \mathcal{L}$ *of* $(\mathcal{P}_o)$ *is characterized by the existence of some* $\bar\omega \in \mathcal{X}^*$ *such that*

$$\begin{cases}
\text{(a)} \quad T\hat\ell \in C,\\
\text{(b)} \quad \big\langle T^*\bar\omega, \hat\ell\big\rangle \leqslant \big\langle T^*\bar\omega, \ell\big\rangle \quad \textit{for all } \ell \in \{\ell \in \mathcal{L};\ T\ell \in C\} \cap \operatorname{dom}\Phi^*,\\
\text{(c)} \quad \hat\ell \in \partial_{\mathcal{L}}\bar\Phi\big(T^*\bar\omega\big).
\end{cases} \tag{5.5}$$

Moreover, $\hat\ell \in \mathcal{L}$ *and* $\bar\omega \in \mathcal{X}^*$ *satisfy* (5.5) *if and only if* $\hat\ell$ *solves* $(\mathcal{P}_o)$ *and* $\bar\omega$ *solves* $(\overline{\mathcal{D}})$.

The assumption $C \cap \operatorname{icor}(T_o \operatorname{dom}\Phi^*) \neq \emptyset$ is equivalent to $C \cap \operatorname{icordom}\Lambda_o^* \neq \emptyset$ and the representation formula (5.5(c)) is equivalent to Young's identity

$$\Phi^*(\hat\ell) + \bar\Phi\big(T^*\bar\omega\big) = \langle\bar\omega, T\hat\ell\rangle = \Lambda^*(\hat x) + \overline{\Lambda}(\bar\omega). \tag{5.6}$$

Formula (5.5(c)) can be made a little more precise by means of the following regularity result.

Theorem 5.7. *Any solution $\bar{\omega}$ of $(\overline{\mathcal{D}})$ shares the following properties*:

(a) $\bar{\omega}$ *is in the* $\sigma(\mathcal{X}^*, \mathcal{X})$*-closure of* $\operatorname{dom}\Lambda$;
(b) $T^*\bar{\omega}$ *is in the* $\sigma(\mathcal{L}^*, \mathcal{L})$*-closure of* $T^*(\operatorname{dom}\Lambda)$.

If in addition the level sets of Φ are $|\cdot|_\Phi$-bounded, then

(a′) $\bar{\omega}$ *is in* $\mathcal{Y}''$. *More precisely, it is in the* $\sigma(\mathcal{Y}'', \mathcal{X})$*-closure of* $\operatorname{dom}\Lambda$;
(b′) $T^*\bar{\omega}$ *is in* $\mathcal{U}''$. *More precisely, it is in the* $\sigma(\mathcal{U}'', \mathcal{L})$*-closure of* $T^*(\operatorname{dom}\Lambda)$,

where $\mathcal{Y}''$ and $\mathcal{U}''$ are the topological *bidual spaces of $\mathcal{Y}$ and $\mathcal{U}$. This occurs if Φ, or equivalently Φ^*, is an even function.*

5.4. Convex conjugates in a Riesz space

The following results are taken from [22,24]. For the basic definitions and properties of Riesz spaces, see [8, Chapter 2].

Let U be a Riesz vector space for the order relation $\leqslant$. Since U is a Riesz space, any $u \in U$ admits a nonnegative part: $u_+ := u \vee 0$, and a nonpositive part: $u_- := (-u) \vee 0$. Of course, $u = u_+ - u_-$ and as usual, we state: $|u| = u_+ + u_-$.

Let Φ be a $[0,\infty]$-valued function on U which satisfies the following conditions:

$$\forall u \in U, \quad \Phi(u) = \Phi(u_+ - u_-) = \Phi(u_+) + \Phi(-u_-), \tag{5.8}$$

$$\forall u, v \in U, \quad \begin{cases} 0 \leqslant u \leqslant v \implies \Phi(u) \leqslant \Phi(v), \\ u \leqslant v \leqslant 0 \implies \Phi(u) \geqslant \Phi(v). \end{cases} \tag{5.9}$$

Clearly (5.8) implies $\Phi(0) = 0$, (5.8) and (5.9) imply that for any $u \in U$, $\Phi(u) = \Phi(u_+) + \Phi(-u_-) \geqslant \Phi(0) + \Phi(0) = 0$. Therefore, Φ^* is $[0,\infty]$-valued and $\Phi^*(0) = 0$.

For all $u \in U$, $\Phi_+(u) = \Phi(|u|)$, $\Phi_-(u) = \Phi(-|u|)$. The convex conjugates of Φ, Φ_+ and Φ_- with respect to $\langle U, U^*\rangle$ are denoted Φ^*, Φ^*_+ and Φ^*_-. Let L be the vector space spanned by $\operatorname{dom}\Phi^*$. The convex conjugates of Φ^*, Φ^*_+ and Φ^*_- with respect to $\langle L, L^*\rangle$ are denoted $\bar{\Phi}$, $\overline{\Phi_+}$ and $\overline{\Phi_-}$. The space of relatively bounded linear forms on U and L are denoted by U^b and L^b, whenever L is a Riesz space.

One writes $a_\pm \in A_\pm$ for $[a_+ \in A_+$ and $a_- \in A_-]$.

Proposition 5.10. *Assume* (5.8) *and* (5.9) *and suppose that L is a Riesz space.*

(a) *For all $\ell \in U^*$,*

$$\Phi^*(\ell) = \begin{cases} \Phi^*_+(\ell_+) + \Phi^*_-(\ell_-) & \text{if } \ell \in U^b, \\ +\infty & \text{otherwise}. \end{cases}$$

(b) *Denoting L_+ and L_- the vector subspaces of L spanned by* $\operatorname{dom}\Phi^*_+$ *and* $\operatorname{dom}\Phi^*_-$, *we have*

$$\bar{\Phi}(\zeta) = \begin{cases} \overline{\Phi_+}(\zeta_{+|L_+}) + \overline{\Phi_-}(\zeta_{-|L_-}) & \text{if } \zeta \in L^b, \\ +\infty & \text{otherwise}, \end{cases}$$

which means that $\overline{\Phi_\pm}(\zeta_\pm) = \overline{\Phi_\pm}(\zeta'_\pm)$ if $\zeta_\pm$ and $\zeta'_\pm$ match on $L_\pm$.

(c) *Let $\ell \in L$, $\zeta \in L^*$ be such that $\ell \in \partial_L \bar{\Phi}(\zeta)$. Then, $\ell_\pm \in \partial_{L_\pm}\overline{\Phi_\pm}(\zeta_{\pm|L_\pm}) \subset L_\pm$.*

Proof. (a) and (b) are proved at [22, Proposition 4.4] under the additional assumption that for all $u \in U$ there exists $\lambda > 0$ such that $\Phi(\lambda u) < +\infty$. But it can be removed. Indeed, if for instance Φ_- is null, Φ^*_- is the convex indicator of $\{0\}$ whose domain is in U^b. The statement about $\bar{\Phi}$ is an iteration of this argument.

The last statement of (b) about $\zeta_{\pm|L_\pm}$ directly follows from $\operatorname{dom}\Phi^*_\pm \subset L_\pm$.

For (c), see the proof of [24, Proposition 4.5]. □

6. Proofs of Theorems 3.2 and 4.2

Theorem 3.2 is a direct corollary of Theorem 4.2.

Proof of Theorem 3.2. Let us assume for the moment that Theorem 4.2 is proved. If

$$\lim_{|t|\to\infty} \gamma^*_z(t)/|t| = \infty, \tag{6.1}$$

for R-a.e. $z \in \mathcal{Z}$, λ is a finite function and Proposition 2.5 insures that whenever $(\mathrm{A}^\forall_\theta)$ holds $T_o^{-1}C \cap L'_{\lambda_\diamond} = \bigcap_{y\in Y}\{\ell \in L'_{\lambda_\diamond};\ \langle\theta,\ell\rangle \geqslant a_y\} = \bigcap_{y\in Y}\{fR \in L_{\lambda^*_\diamond}R;\ \int_{\mathcal{Z}}\langle y,\theta\rangle f\,dR \geqslant a_y\}$. In particular the problems (P_C) and $(\bar{\mathrm{P}}_C)$ are equivalent.

On the other hand, one shows as at Remark 3.5(a) that $\theta(z)=0$ for any z for which (6.1) fails. Hence, the problems (P_C) and $(\bar{\mathrm{P}}_C)$ are equivalent under $(\mathrm{A}_\theta^\forall)$ and Theorem 3.2 follows from Theorem 4.2. □

The remainder of this section is devoted to the proof of Theorem 4.2. It is an application of Theorems 5.3 and 5.4.

Lemma 6.2. *Assume* (A_R) *and* $(\mathrm{A}^1_{\gamma^*})$. *We have*

$$I(fR)=\sup_{u\in L_{\lambda_\diamond}}\big\{\langle u,f\rangle_{L_{\lambda_\diamond},L_{\lambda_\diamond^*}}-I_\gamma(u)\big\},\quad f\in L_{\lambda_\diamond^*},\tag{6.3}$$

$$\bar I(\ell)=\sup_{u\in L_{\lambda_\diamond}}\big\{\langle u,\ell\rangle_{L_{\lambda_\diamond},L'_{\lambda_\diamond}}-I_\gamma(u)\big\},\quad \ell\in L'_{\lambda_\diamond}.\tag{6.4}$$

Moreover, I is $\sigma(L_{\lambda_\diamond^*}R,L_{\lambda_\diamond})$-lower semicontinuous and $\bar I$ is $\sigma(L'_{\lambda_\diamond},L_{\lambda_\diamond})$-inf-compact.

If in addition (6.1) *holds for R-a.e. $z\in\mathcal{Z}$,*

$$I(fR)=\sup_{u\in E_{\lambda_\diamond}}\big\{\langle u,f\rangle_{L_{\lambda_\diamond},L_{\lambda_\diamond^*}}-I_\gamma(u)\big\},\quad f\in L_{\lambda_\diamond^*},\tag{6.5}$$

and I is $\sigma(L_{\lambda_\diamond^*}R,E_{\lambda_\diamond})$-inf-compact.

In other words, (I,I_γ) and $(\bar I,I_\gamma)$ are respectively convex conjugate to each other for the dualities $\langle L_{\lambda_\diamond^*}R,L_{\lambda_\diamond}\rangle$ and $\langle L'_{\lambda_\diamond},L_{\lambda_\diamond}\rangle$. It follows that $\bar I$ is the greatest convex $\sigma(L'_{\lambda_\diamond},L_{\lambda_\diamond})$-lower semicontinuous extension of I from $L_{\lambda_\diamond^*}R$ to $L'_{\lambda_\diamond}$.

Proof. Under the assumptions (A_R) and $(\mathrm{A}^1_{\gamma^*})$, γ^* is a normal convex integrand. It follows that I_{γ^*} on $L_{\lambda_\diamond^*}$ and I_γ on $L_{\lambda_\diamond}$ are convex conjugate to each other, see [26], which is (6.3). One obtains (6.5) similarly.

Let us prove (6.4). Fix $\ell\in L'_{\lambda_\diamond}$. Clearly, $\sup_{u\in L_{\lambda_\diamond}}\{\langle u,\ell\rangle-I_\gamma(u)\}=\sup_{u\in L_{\lambda_\diamond}}\{\langle u,\ell^a+\ell^s\rangle-I_\gamma(u)-\iota_{\mathrm{dom}\,I_\gamma}(u)\}\leqslant \sup_{u\in L_{\lambda_\diamond}}\{\langle u,\ell^a\rangle-I_\gamma(u)\}+\sup_{u\in L_{\lambda_\diamond}}\{\langle u,\ell^s\rangle-\iota_{\mathrm{dom}\,I_\gamma}(u)\}=\bar I(\ell)$.

Let us prove the converse inequality. For each $\epsilon>0$ there exists a measurable set S_ϵ such that $R(S_\epsilon)\leqslant\epsilon$ and $\langle u\mathbf{1}_{S_\epsilon^c},\ell^s\rangle=0$ for all $u\in L_{\lambda_\diamond}$. Hence,

$$\begin{aligned}\sup_{u\in L_{\lambda_\diamond}}\big\{\langle u,\ell\rangle-I_\gamma(u)\big\}&=\sup_{u,v\in L_{\lambda_\diamond}}\big\{\big\langle u\mathbf{1}_{S_\epsilon^c}+v\mathbf{1}_{S_\epsilon},\ell^a+\ell^s\big\rangle-I_\gamma(u\mathbf{1}_{S_\epsilon^c}+v\mathbf{1}_{S_\epsilon})\big\}\\&=I\big(\mathbf{1}_{S_\epsilon^c}\ell^a\big)+\sup_{v\in L_{\lambda_\diamond}}\big\{\big\langle v\mathbf{1}_{S_\epsilon},\ell^a\big\rangle-I_\gamma(v\mathbf{1}_{S_\epsilon})+\big\langle v,\ell^s\big\rangle-\iota_{\mathrm{dom}\,I_\gamma}(v\mathbf{1}_{S_\epsilon})\big\}.\end{aligned}$$

Taking $v_\delta\in L_{\lambda_\diamond}$ such that $v_\delta\mathbf{1}_{S_\epsilon}\in\mathrm{dom}\,I_\gamma$ and $\langle v_\delta,\ell^s\rangle\geqslant\min(\iota^*_{\mathrm{dom}\,I_\gamma}(\ell^s)-\delta,1/\delta)$ with $\delta>0$, one obtains $\sup_{u\in L_{\lambda_\diamond}}\{\langle u,\ell\rangle-I_\gamma(u)\}\geqslant\min(I(\mathbf{1}_{S_\epsilon^c}\ell^a)+\iota^*_{\mathrm{dom}\,I_\gamma}(\ell^s)+\langle v_\delta\mathbf{1}_{S_\epsilon},\ell^a\rangle-I_\gamma(v_\delta\mathbf{1}_{S_\epsilon})-\delta,1/\delta)$. By dominated convergence, letting ϵ tend to zero, we see that $\sup_{u\in L_{\lambda_\diamond}}\{\langle u,\ell\rangle-I_\gamma(u)\}\geqslant\min(I(\ell^a)+\iota^*_{\mathrm{dom}\,I_\gamma}(\ell^s)-\delta,1/\delta)$ for any $\delta>0$. As $\delta>0$ is arbitrary, $\sup_{u\in L_{\lambda_\diamond}}\{\langle u,\ell\rangle-I_\gamma(u)\}\geqslant\bar I(\ell)$. This completes the proof of (6.4).

Because of these variational representations, I and $\bar I$ are lower semicontinuous with respect to the corresponding weak topologies and the compactness of their level sets is a standard consequence of Banach–Alaoglu theorem and the strong continuity of I_γ, see [22, Corollary 2.2] for instance.

Note that assuming (6.1) in the last statement is necessary for $E_{\lambda_\diamond}$ not to be reduced to the null space. □

Proof of Theorem 4.2. One applies the abstract results of Section 5.3 with

$$\Phi(u)=I_\lambda(u):=\int_{\mathcal{Z}}\lambda(u)\,dR,\quad u\in\mathcal{U}_o:=\mathcal{L}_{\lambda_\diamond}.\tag{6.6}$$

This gives $\mathcal{U}=L_{\lambda_\diamond}$ with the Orlicz norm $|u|_\Phi=\|u\|_{\lambda_\diamond}$ and $\mathcal{L}=L'_{\lambda_\diamond}=L_{\lambda_\diamond^*}R\oplus L^s_{\lambda_\diamond}$, by Theorem 2.3. The spaces $\mathcal{Y}$ and $\mathcal{X}$ match with the definitions of Section 2.5.

- *Reduction to* $m=0$. We have seen at (2.10) that the transformation $Q\rightsquigarrow\ell=Q-mR$ corresponds to the transformations $\gamma\rightsquigarrow\lambda$ and $(\mathrm{P}_C)\rightsquigarrow$ (2.10). This still works with $(\bar{\mathrm{P}}_C)$ and one can assume from now on without loss of generality that under $(\mathrm{A}^2_{\gamma^*})$, $m=0$ and $\gamma=\lambda$.

  The hypothesis $(\mathrm{A}^2_{\gamma^*})$ will not be used during the rest of the proof. This allows Remarks 3.5(b) and 4.6(d).
- *Verification of* (H_C). It is equivalent to (4.3).
- *Verification of* (H_T). The hypothesis (H_{T1}) is $(\mathrm{A}^\exists_\theta)$ while (H_{T2}) is (A^2_θ).
- *Verification of* (H_Φ). Suppose that $W=\{z\in\mathcal{Z};\ \lambda(z,s)=0,\ \forall s\in\mathbb{R}\}$ is such that $R(W)>0$. Then, any ℓ such that $\langle u\mathbf{1}_W,\ell\rangle>0$ for some $u\in L_{\lambda_\diamond}$ satisfies $\Phi^*(\ell)=+\infty$. Therefore, one can remove W from $\mathcal{Z}$ without loss of generality. Once, this is done, the hypothesis (H_Φ) is satisfied under the hypothesis $(\mathrm{A}^1_{\gamma^*})$.

With (5.2), this implies that

$$TL'_{\lambda_\diamond} \subset \mathcal{X} \tag{6.7}$$

and by Theorem 5.3(a)

$$\operatorname{dom}\Gamma^* \subset \mathcal{X}. \tag{6.8}$$

With (6.6) and Theorem 5.3(a), $\operatorname{dom}\Phi^* \subset \mathcal{L} = L'_{\lambda_\diamond}$. Hence, with Lemma 6.2 one obtains

$$\bar{I}(\ell) = \Phi^*(\ell), \quad \ell \in L'_{\lambda_\diamond}. \tag{6.9}$$

- *The computation of $\bar{\Phi}$ in the case where λ is even.* Since Φ is even, Theorem 5.7 tells us that $\operatorname{dom}\bar{\Phi}$ is included in the $\sigma(L''_\lambda, L'_\lambda)$-closure of $\operatorname{dom}\Phi$. Thanks to the decomposition (4.1) and Lemma 6.2 applied with λ^* instead of γ, the extension $\bar{\Phi}$ is given for each $\zeta \in L''_\lambda$ by

$$\begin{aligned}\bar{\Phi}(\zeta) &= (\bar{I}_{\lambda^*})^*(\zeta_1, \zeta_2) = I^*_{\lambda^*}(\zeta_1) + \iota^{**}_{\operatorname{dom} I_\lambda}(\zeta_2)\\ &= \bar{I}_\lambda(\zeta_1) + \iota_D(\zeta_2) = I_\lambda\big(\zeta_1^a\big) + \iota^*_{\operatorname{dom} I_{\lambda^*}}\big(\zeta_1^s\big) + \iota_D(\zeta_2),\end{aligned}$$

where D is the $\sigma(L^{s\prime}_\lambda, L^s_\lambda)$-closure of $\operatorname{dom} I_\lambda$.
- *Extension to the case where λ is not even.* By Proposition 5.10(b), we have $\bar{\Phi}(\zeta) = \bar{\Phi}_+(\zeta_{+|L'_{\lambda_+}\cap L'_{\lambda_\diamond}}) + \bar{\Phi}_-(\zeta_{-|L'_{\lambda_-}\cap L'_{\lambda_\diamond}})$ if $\zeta \in L'^{b}_{\lambda_\diamond}$ and $+\infty$ otherwise. It follows that

$$\bar{\Phi}(\zeta) = \bar{I}^*_\lambda(\zeta) = I_\lambda\big(\zeta_1^a\big) + \iota^*_{\operatorname{dom} I_{\lambda^*}}\big(\zeta_1^s\big) + \iota_D(\zeta_2) \tag{6.10}$$

if $\zeta \in K''_\lambda$ and $+\infty$ otherwise. In particular, we have

$$\Lambda(y) = I_\lambda\big(\langle y, \theta\rangle\big), \quad y \in \mathcal{Y},$$
$$\overline{\Lambda}(\omega) = \begin{cases} I_\lambda([T^*\omega]_1^a) + \iota^*_{\operatorname{dom} I_{\lambda^*}}([T^*\omega]_1^s) + \iota_D([T^*\omega]_2) & \text{if } \omega \in \widehat{\mathcal{Y}},\\ +\infty & \text{otherwise,}\end{cases} \quad \omega \in \mathcal{X}^*,$$

where $\widehat{\mathcal{Y}} \subset \mathcal{X}^*$ is the extension of $\mathcal{Y}_o$ which is isomorphic to the $\sigma(K''_\lambda, L'_\pm)$-closure $\overline{T_o^*\mathcal{Y}_o}$ of $T_o^*\mathcal{Y}_o$ in the sense that $T^*\overline{\mathcal{Y}} = \overline{T_o^*\mathcal{Y}_o}$. Considering the maximization problem

$$\textit{maximize} \ \inf_{x \in C\cap\mathcal{X}} \langle \omega, x\rangle - I_\lambda\big(\big[T^*\omega\big]_1^a\big) - \iota^*_{\operatorname{dom} I_{\lambda^*}}\big(\big[T^*\omega\big]_1^s\big) - \iota_D\big(\big[T^*\omega\big]_2\big), \quad \omega \in \widehat{\mathcal{Y}}, \tag{6.11}$$

this provides us with the dual problems $(\mathcal{D}) = (\mathrm{D}_C)$ and $(\overline{\mathcal{D}}) = (6.11)$.
- *Proof of* (a) *and* (b)*.* Apply Theorem 5.3.
Let us go on with the proof of (c). By Theorem 5.4, ($(\bar{\mathrm{P}}_C)$, (6.11)) admits a solution in $L'_{\lambda_\diamond} \times \widehat{\mathcal{Y}}$ and $(\hat{\ell}, \bar{\omega}) \in L'_{\lambda_\diamond} \times \widehat{\mathcal{Y}}$ solves ($(\bar{\mathrm{P}}_C)$, (6.11)) if and only if

$$\begin{cases} \text{(a)} & \hat{x} \in C \cap \operatorname{dom}\Gamma^*,\\ \text{(b)} & \langle \bar{\omega}, \hat{x}\rangle \leqslant \langle \bar{\omega}, x\rangle, \quad \forall x \in C \cap \operatorname{dom}\Gamma^*,\\ \text{(c)} & \hat{\ell} \in \partial_{L'_{\lambda_\diamond}} \bar{\Phi}\big(T^*\bar{\omega}\big),\end{cases} \tag{6.12}$$

where $\hat{x} \triangleq T\hat{\ell}$ is defined in the weak sense with respect to the duality $\langle \mathcal{Y}, \mathcal{X}\rangle$. Since $\operatorname{dom}\Gamma^* \subset \mathcal{X}$, the above dual brackets are meaningful.
- *The computation of* $\partial_{L'_\lambda}\bar{\Phi}(\zeta)$. Let us first assume that λ is even. For all $u \in L_\lambda$, $u_1^a = u_2 = u$ and $u_1^s = 0$. This gives $\bar{\Phi}(\zeta + u) - \bar{\Phi}(\zeta) = I_\lambda(\zeta_1^a + u_1) - I_\lambda(\zeta_1^a) + \iota_D(\zeta_2 + u_2) - \iota_D(\zeta_2)$ where $u_1 = u$ and $u_2 = u$ act respectively on $L_{\lambda^*}R$ and L^s_λ. This direct sum structure leads us to

$$\partial_{L'_\lambda}\bar{\Phi}(\zeta) = \partial_{L_{\lambda^*}R} I_\lambda\big(\zeta_1^a\big) + \partial_{L^s_\lambda}\iota_D(\zeta_2), \tag{6.13}$$

which again is the direct sum of the absolutely continuous and singular components of $\partial_{L'_\lambda}\bar{\Phi}(\zeta)$. Differentiating in the directions of $\mathcal{U} = L_\lambda$, one obtains $\partial_{L_{\lambda^*}R} I_\lambda(\zeta_1^a) = \{\lambda'(\zeta_1^a)R\}$. The computation of $\partial_{L^s_\lambda}\iota_D(\zeta_2)$ is standard: $\partial_{L^s_\lambda}\iota_D(\zeta_2) = D^\perp(\zeta_2)$ is the outer normal cone of D at ζ_2.
Now, consider a general λ. By Proposition 5.10(a), $\hat{\ell}_+ \in \partial_{L'_{\lambda_+}}\bar{\Phi}_+([T^*\bar{\omega}]_+)$ and $\hat{\ell}_- \in \partial_{L'_{\lambda_-}}\bar{\Phi}_-([T^*\bar{\omega}]_-)$. Therefore, (6.13) becomes

$$\partial_{L'_{\lambda_\diamond}}\bar{\Phi}(\zeta) = \partial_{K_{\lambda^*}R} I_\lambda\big(\zeta_1^a\big) + \partial_{K^s_\lambda}\iota_D(\zeta_2).$$

- *From* (6.11) *to* ($\bar{\mathrm{D}}_C$). The solution $(\hat{\ell}, \bar{\omega}) \in L'_{\lambda_\diamond} \times \widehat{\mathcal{Y}}$ of (($\bar{\mathrm{P}}_C$), (6.11)) satisfies $\hat{\ell}^a \in \partial_{L_{\lambda_\diamond^*} R} \bar{\Phi}_a([T^*\bar{\omega}]_1)$ where $\bar{\Phi}_a([T^*\omega]_1) = I_\lambda([T^*\omega]_1^a) + \iota^*_{\operatorname{dom} I_{\lambda^*}}([T^*\omega]_1^s)$. This implies that $[T^*\bar{\omega}]_1^s \in \partial \iota^{**}_{\operatorname{dom} I_{\lambda^*}}(\hat{\ell}^a)$. But under the assumption (4.4), $\hat{\ell}^a$ is an interior point so that
$$\big[T^*\bar{\omega}\big]_1^s = 0. \tag{6.14}$$
Now, plug the identity $[T^*\omega]_1^s = 0$ into (6.11) and note that maximizing $-\iota_D$ amounts to restrict the problem to D to see that one can replace (6.11) with ($\bar{\mathrm{D}}_C$) under the assumption (4.4).
- *Representation of* $[T^*\bar{\omega}]_1^a$. One still has to prove that
$$\big[T^*\bar{\omega}\big]_1^a(z) = \big\langle \theta(z), \tilde{\omega}\big\rangle \tag{6.15}$$
for R-a.e. $z \in \mathcal{Z}$ and some linear form $\tilde{\omega}$ on $\mathcal{X}_o$.
If $W_- := \{z \in \mathcal{Z};\ \lambda(z,s) = 0,\ \forall s \leqslant 0\}$ satisfies $R(W_-) > 0$, $\operatorname{dom} \bar{I}$ is a set of linear forms which are nonnegative on W and $\gamma_z'(s) = 0$ for all $s \leqslant 0, z \in W$. Hence, one can take any function for the restriction of $[T^*\bar{\omega}]_{1-}^a$ to W_- without modifying (6.12)(c). As a symmetric remark holds for $W_+ = \{z \in \mathcal{Z};\ \lambda(z,s) = 0,\ \forall s \geqslant 0\}$, it remains to consider the situation where for R-a.e. z, there are $s_-(z) < 0 < s_+(z)$ such that $\lambda(z, s_\pm(z)) > 0$. This implies that $\lim_{s \to \pm\infty} \lambda(z,s)/s > 0$.
By Theorem 5.4, $T^*\bar{\omega}$ is in the $\sigma(K''_\lambda, L'_\pm)$-closure of $T^*(\operatorname{dom} \Lambda)$. Therefore, $[T^*\bar{\omega}]_1^a$ is in the $\sigma(K_\lambda, L'_\pm)$-closure of $T^*(\operatorname{dom} \Lambda)$. As $T^*(\operatorname{dom} \Lambda)$ is convex, this closure is its strong closure in K_λ. Since there exists a finite measurable function $c(z)$ such that $0 < c(z) \leqslant \lim_{s \to \infty} \lambda(z,s)/s$, one can consider the nontrivial Young function $\rho(z,s) = c(z)|s|$ and the corresponding Orlicz spaces L_ρ and $L'_\rho = L_{\rho^*}$. If R is a bounded measure, we have $L_{\lambda_\diamond} \subset L_\rho$ and $L_{\rho^*} \subset L_{\lambda_\diamond^*}$, so that $[T^*\bar{\omega}]_1^a$ is in the strong closure of $T^*(\operatorname{dom} \Lambda)$ in L_ρ.
As a consequence, $[T^*\bar{\omega}]_1^a$ is the pointwise limit of a sequence $(T^* y_n)_{n \geqslant 1}$ with $y_n \in \mathcal{Y}$. As $T^* y_n(z) = \langle y_n, \theta(z)\rangle$, we see that $[T^*\bar{\omega}]_1^a(z) = \langle \theta(z), \tilde{\omega}\rangle$ for some linear form $\tilde{\omega}$ on $\mathcal{X}_o$. If R is unbounded, it is still assumed to be σ-finite: there exists a sequence $(\mathcal{Z}_k)$ of measurable subsets of $\mathcal{Z}$ such that $\bigcup_k \mathcal{Z}_k = \mathcal{Z}$ and $R(\mathcal{Z}_k) < \infty$ for each k. Hence, for each k and all $z \in \mathcal{Z}_k$, $(T^*\bar{\omega})^a(z) = \langle \theta(z), \tilde{\omega}^k\rangle$ for some linear form $\tilde{\omega}^k$ on $\mathcal{X}_o$, from which (6.15) follows.
- *Proof of* (c). It follows from the previous considerations and Theorem 5.4.
- *Proof of* (d). Statement (d)1 follows from Theorem 5.3. Statement (d)2 is immediately deduced from (c). Finally, (d)3 is (5.6). □

7. Examples

Examples of entropy minimization problems are presented.

7.1. Some examples of entropies related to the Boltzmann entropy

The entropies defined below occur naturally in statistical physics, probability theory, mathematical statistics and information theory.

Boltzmann entropy

The Boltzmann entropy with respect to the positive measure R is defined by
$$H_B(Q|R) = \begin{cases} \int_{\mathcal{Z}} \log(\frac{dQ}{dR})\, dQ & \text{if } 0 \leqslant Q \prec R, \\ +\infty & \text{otherwise,} \end{cases}$$
for each $Q \in M_{\mathcal{Z}}$. It corresponds to
$$\gamma_z^*(t) = \begin{cases} t\log t & \text{if } t > 0, \\ 0 & \text{if } t = 0, \\ +\infty & \text{if } t < 0. \end{cases}$$
But this γ^* takes negative values and is ruled out by our hypotheses. A way to circumvent this problem is to consider the variant below.

A variant of the Boltzmann entropy

Let $m : \mathcal{Z} \to (0,\infty)$ be a positive measurable function. Considering
$$\gamma_z^*(t) = t\log t - \big[1 + \log m(z)\big]t + m(z), \quad t > 0,$$
one sees that it is nonnegative and that $\gamma_z^*(t) = 0$ if and only if $t = m(z)$. Hence γ^* enters the framework of this paper and
$$\lambda_z(s) = m(z)\big[e^s - s - 1\big], \quad s \in \mathbb{R}. \tag{7.1}$$
It is easily seen that

$$H_B(Q\,|R) = I_{\gamma^*}(Q) + \int_{\mathcal{Z}} (1+\log m)\,dQ - \int_{\mathcal{Z}} m\,dR$$

which is meaningful if Q integrates $1+\log m$ where $m \in L_1(R)$. As an application, let R be the Lebesgue measure on $\mathcal{Z} = \mathbb{R}^d$ and minimize $H_B(Q\,|R)$ on the set $\mathcal{C} = \{Q \in P_{\mathcal{Z}};\ \int_{\mathcal{Z}} |z|^2 Q(dz) = E\} \cap \mathcal{C}_o$. Taking $m(z) = e^{-|z|^2}$, one is led to minimizing I_{γ^*} on $\mathcal{C}$.

A special case

It is defined by

$$H(Q\,|R) = \begin{cases} \int_{\mathcal{Z}} [\frac{dQ}{dR} \log(\frac{dQ}{dR}) - \frac{dQ}{dR} + 1]\,dR & \text{if } 0 \leqslant Q \prec R, \\ +\infty & \text{otherwise,} \end{cases} \qquad Q \in M_{\mathcal{Z}}. \tag{7.2}$$

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 $\lambda_z(s) = e^s - s - 1$, $s \in \mathbb{R}$ for all $z \in \mathcal{Z}$. Note that $H(Q\,|R) < \infty$ implies that Q is nonnegative.

Relative entropy

The reference measure R is assumed to be a probability measure and one denotes $P_{\mathcal{Z}}$ the set of all probability measures on $\mathcal{Z}$. The relative entropy of $Q \in M_{\mathcal{Z}}$ with respect to $R \in P_{\mathcal{Z}}$ is the following variant of the Boltzmann entropy:

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

It is (7.2) with the additional constraint that $Q(\mathcal{Z}) = 1$:

$$I(Q\,|R) = H(Q\,|R) + \iota_{\{Q(\mathcal{Z})=1\}}.$$

When minimizing the Boltzmann entropy $Q \mapsto H_B(Q\,|R)$ on a constraint set which is included in $P_{\mathcal{Z}}$, we have for all $P, Q \in P_{\mathcal{Z}}$,

$$H_B(Q\,|R) = I(Q\,|P) + \int_{\mathcal{Z}} \log\left(\frac{dP}{dR}\right) dQ$$

which is meaningful for each $Q \in P_{\mathcal{Z}}$ which integrates $\frac{dP}{dR}$.

Extended relative entropy

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

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

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

The extended relative entropy is defined by

$$\bar{I}(\ell|R) = I(\ell^a|R) + \sup\left\{ \langle \ell^s, u\rangle;\ u, \int_{\mathcal{Z}} e^u\,dR < \infty \right\}, \quad \ell \in \mathcal{E}(\mathcal{Z}), \tag{7.3}$$

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

Reverse relative entropy

The reference measure R is assumed to be a probability measure. The reverse relative entropy is

$$Q \in M_{\mathcal{Z}} \mapsto \begin{cases} I(R|Q) & \text{if } Q \in P_{\mathcal{Z}}, \\ +\infty & \text{otherwise} \end{cases} \in [0,\infty].$$

It corresponds to

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

$$\lambda_z(s) = \begin{cases} -\log(1-s) - s & \text{if } s < 1, \\ +\infty & \text{if } s \geqslant 1, \end{cases} \tag{7.4}$$

for all $z \in \mathcal{Z}$, with the additional constraint that $Q(\mathcal{Z}) = 1$.

7.2. Some examples of constraints

Let us consider the three standard constraints which are the moment, marginal and Fredholm constraints. We are going to give a precise formulation of (1.7)–(1.9) by specifying the adjoint operator T_o^* at (7.5), (7.6) and (7.9).

Moment constraints

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

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

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

$$T_o^* y(z) = \sum_{1\leqslant k\leqslant K} y_k\theta_k(z), \quad y = (y_1,\dots,y_K) \in \mathbb{R}^K, \ z \in \mathcal{Z}. \tag{7.5}$$

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, $(\ell_A, \ell_B) = \int_{A\times B} (\delta_a, \delta_b)\ell(da\,db)$.

More precisely, let U_A, U_B be the spaces of bounded 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 $\mathcal{X}_o = U_A^* \times U_B^*$. It is easy to see that the adjoint of the marginal operator

$$T_o\ell = (\ell_A, \ell_B) \in U_A^* \times U_B^*, \quad \ell \in \mathcal{L}_o = U_{AB}^*,$$

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

$$T_o^*(f,g) = f \oplus g \in U_{AB}, \quad f \in U_A, \ g \in U_B, \tag{7.6}$$

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

Generalized Fredholm integral constraints

In addition to $(\mathcal{Z}, R)$ and the function γ, we consider a measure space $(\mathcal{R}, \rho)$ where ρ is a nonnegative measure on $\mathcal{R}$ and a pair (X, Y) of vector spaces in separating duality. Let $\theta : \mathcal{R} \times \mathcal{Z} \to X$ be an X-valued function on $\mathcal{R} \times \mathcal{Z}$ and $x = (x_r)_{r\in\mathcal{R}}$ be an element of $X^{\mathcal{R}}$. We are going to give some meaning at (7.9) to the following formal expression of the Fredholm integral constraint

$$\big\langle \theta(r,\cdot), \ell \big\rangle = x_r \quad \text{for } \rho\text{-almost every } r \in \mathcal{R}, \ \ell \in L'_{\lambda_\diamond}. \tag{7.7}$$

To do this, let us consider a vector space $\mathcal{Y}_o$ of functions $y:\mathcal{R}\to Y$ and a vector space $\mathcal{X}_o$ of functions $x:\mathcal{R}\to X$. Denote $y=(y_r)_{r\in\mathcal{R}}\in\mathcal{Y}_o\subset Y^{\mathcal{R}}$, $x=(x_r)_{r\in\mathcal{R}}\in\mathcal{X}_o\subset X^{\mathcal{R}}$ and assume that for any $x\in\mathcal{X}_o$ and $y\in\mathcal{Y}_o$,

$$r\in\mathcal{R}\mapsto\langle y_r,x_r\rangle\in\mathbb{R}\text{ is measurable}\quad\text{and}\quad\int_{\mathcal{R}}\big|\langle y_r,x_r\rangle\big|\,\rho(dr)<\infty. \tag{7.8}$$

In other words $\langle y_\cdot,x_\cdot\rangle\in L_1(\mathcal{R},\rho)$. This property allows us to define the dual bracket between $\mathcal{Y}_o$ and $\mathcal{X}_o$: $\langle y,x\rangle_{\mathcal{Y}_o,\mathcal{X}_o}=\int_{\mathcal{R}}\langle y_r,x_r\rangle\,\rho(dr)$. One identifies $\mathcal{X}_o$ as a subset of $\mathcal{Y}_o^*$. In order that $\mathcal{Y}_o$ separates $\mathcal{X}_o$, $x,x'\in\mathcal{X}_o$ are identified if: $\langle y,x-x'\rangle_{\mathcal{Y}_o,\mathcal{X}_o}=0$, $\forall y\in\mathcal{Y}_o$. We assume that the function $\theta:\mathcal{R}\times\mathcal{Z}\to X$ appearing in (7.7) satisfies the following conditions:

(1) for any $z\in\mathcal{Z}$, the function $\theta(\cdot,z):r\mapsto\theta(r,z)$ belongs to $\mathcal{X}_o$;
(2) for any $y\in\mathcal{Y}_o$, the function $\langle y,\theta(\cdot)\rangle_{\mathcal{Y}_o,\mathcal{X}_o}:z\mapsto\int_{\mathcal{R}}\langle y_r,\theta(r,z)\rangle\,\rho(dr)$ is measurable;
(3) for any $y\in\mathcal{Y}_o$, there exists $\alpha>0$ such that $\int_{\mathcal{Z}}\lambda_\diamond(\alpha\langle y,\theta(\cdot)\rangle_{\mathcal{Y}_o,\mathcal{X}_o})\,dR<\infty$.

In other words, for all $y\in\mathcal{Y}_o$, $\langle y,\theta(\cdot)\rangle_{\mathcal{Y}_o,\mathcal{X}_o}\in L_{\lambda_\diamond}$. Finally, if $\mathcal{Y}_o$ is a rich enough space in the sense that $\int_{\mathcal{R}}\langle y_r,x_r\rangle\,\rho(dr)=0$ for all $y\in\mathcal{Y}_o$ implies that $x_r=0$ for ρ-a.e. r, one can reformulate correctly (7.7) by

$$\forall y\in\mathcal{Y}_o,\quad \big\langle\ell,\langle y,\theta(\cdot)\rangle_{\mathcal{Y}_o,\mathcal{X}_o}\big\rangle_{L'_{\lambda_\diamond},L_{\lambda_\diamond}}=\int_{\mathcal{R}}\langle y_r,x_r\rangle_{Y,X}\,\rho(dr).$$

This corresponds to

$$T_o^*y(z)=\int_{\mathcal{R}}\big\langle y_r,\theta_r(z)\big\rangle_{Y,X}\,\rho(dr),\quad y\in\mathcal{Y}_o,\ z\in\mathcal{Z}. \tag{7.9}$$

The moment constraints (7.5) are recovered choosing $\mathcal{R}=\{1,\dots,K\}$ with the counting measure $\rho=\sum_{1\leqslant r\leqslant K}\delta_r$ for all $z\in\mathcal{Z}$ and $r\in\mathcal{R}$.

The marginal constraint (7.6) is recovered with $\mathcal{R}=\{0,1\}$, $\rho=\delta_0+\delta_1$, $\theta_0(a,b)=\delta_a$ and $\theta_1(a,b)=\delta_b$ for each $z=(a,b)\in\mathcal{Z}=A\times B$ and $y=(y_0,y_1)\in\mathcal{Y}_o=U_A\times U_B$.

A Fredholm constraint for a random path

Let $\mathcal{Z}=\mathcal{C}([0,1],Z)$ be the set of all continuous paths from $[0,1]$ to the topological state space Z. Take $\Phi:Z\to\mathbb{R}$ a measurable function and consider the constraint $\int_{\mathcal{Z}}\Phi(\zeta_r)\,dP=x_r$ for each $r\in[0,1]$ with $x=(x_r)_{0\leqslant r\leqslant 1}\in\mathbb{R}^{[0,1]}$ where $\zeta_r(z)=z_r$ is the position at time r of the path $z=(z_r)_{0\leqslant r\leqslant 1}$. Applying (7.9), one sees that the constraint

$$\left(\int_{\mathcal{Z}}\Phi(\zeta_r)\,dP\right)_{0\leqslant r\leqslant 1}\in C,\quad P\in P_{\mathcal{C}([0,1],Z)}, \tag{7.10}$$

with C a convex subset of $\mathbb{R}^{[0,1]}$, is determined by

$$T_o^*y(z)=\int_0^1 y_r\Phi(z_r)\,dr,\quad y\in\mathcal{Y}_o,\ z=(z_r)_{0\leqslant r\leqslant 1}\in\mathcal{C}\big([0,1],Z\big), \tag{7.11}$$

with $\mathcal{Y}_o=\mathcal{C}([0,1],\mathbb{R})$ or $\mathcal{Y}_o=\mathcal{S}([0,1])$: the space of simple functions on $[0,1]$.

7.3. Relative entropy under good constraints

We considerer the minimization of the relative entropy $I(\cdot|R)$

$$\textit{minimize } I(P|R)\quad\textit{subject to}\quad\int_{\mathcal{Z}}\theta\,dP\in C,\quad P\in P_{\mathcal{Z}}, \tag{7.12}$$

under marginal constraints and under (7.10).

(a) Since the marginal constraint is bounded, see (7.6), $(\mathrm{A}_\theta^\forall)$ holds. Applying Theorem 3.2 with C a singleton, one recovers the results of [10] and [23] where this problem is solved in details with a different approach. In particular, $\gamma'(s)=e^s$ and by Theorem 3.2(c), the solution of (7.12) is

$$\widehat{P}(da\,db)=f(a)g(b)\,R(da\,db) \tag{7.13}$$

where f, g are functions on A and B such that $f\otimes g$ is R-measurable. In fact, Theorem 3.2 tells us that $\widehat{P}(da\,db)=e^{\tilde\omega(a,b)}R(da\,db)$ with $\tilde\omega$ R-measurable in the $\sigma(K_\tau,K_{\tau^*})$-closure of $\{\varphi\oplus\psi;\ \varphi\in U_A,\ \psi\in U_B\}$ and one can prove that $\tilde\omega=u\oplus v$ for some u and v. Note that it is not stated that $f=e^u$ and $g=e^v$ are measurable. The product form of (7.13) plays an important role in Euclidean quantum mechanics [9,14,28].

(b) Let us have a look at the minimization of the relative entropy with respect to $R \in P_{\mathcal{C}([0,1],Z)}$ under the constraint (7.10). Again, if Φ is bounded, $T_o^* y$ given by (7.11) is bounded for any $y \in \mathcal{Y}_o$ and $(\mathrm{A}_\theta^\forall)$ holds. The solution of (7.12) is

$$\widehat{P}(dz) = \frac{1}{\mathcal{N}} \exp\left(\int_{[0,1]} \tilde{\omega}_r \Phi(z_r)\, dr \right) R(dz)$$

where $\mathcal{N} > 0$ is a normalizing constant and $\tilde{\omega}$ is a measurable function on $[0,1]$. This still holds true if Φ satisfies $\int_{\mathcal{Z}} e^{\alpha \int_{[0,1]} |\Phi(z_r)|\, dr} R(dz) < \infty$ for all $\alpha \geqslant 0$.

7.4. Extended relative entropy under critical constraints

We considerer the minimization of the *extended* relative entropy $\bar{I}(\cdot|R)$ under some simple critical constraints:

$$\textit{minimize } \bar{I}(\ell|R) \quad \textit{subject to} \quad \langle \theta, \ell \rangle \in C, \quad \ell \in \mathcal{E}(\mathcal{Z}). \tag{7.14}$$

(a) Take the probability measure $R(dz) = e^{-z}\, dz$ on the space $\mathcal{Z} = [0,\infty)$ and $\theta(z) = z$, $z \in \mathcal{Z}$. The constraint to be considered is

$$\langle \theta, \ell \rangle \geqslant c, \quad \ell \in \mathcal{E}(\mathcal{Z}), \tag{7.15}$$

for some $c \geqslant \int_{[0,\infty)} \theta\, dR = 1$. In restriction to $P_{\mathcal{Z}}$ this gives $\int_{[0,\infty)} \theta\, dP \geqslant c$, $P \in P_{\mathcal{Z}}$. By (7.5), we have $T_o^* y(z) = yz$ for each real y. This constraint is critical since $\int_{[0,\infty)} e^{yz} R(dz) = 1/(1-y)$ if $y < 1$ and $+\infty$ if $y \geqslant 1$. In other words, θ belongs to $L_\tau(R)$ but not to $E_\tau(R)$.
For each $y < 1$, denote $P_y(dz) := e^{yz - \Lambda(y)} R(dz)$ with the normalizing factor $\Lambda(y) := \log \int_{[0,\infty)} e^{yz} R(dz) = -\log(1-y)$. By Theorem 4.2, the unique solution of (7.14) is $P_{y(c)}$ where $y(c) < 1$ is the solution of the equation

$$\Lambda'(y) = \int_{[0,\infty)} \theta\, dP_y = c \tag{7.16}$$

that is $y(c) = 1 - 1/c$. We have shown that $\widehat{P}(dz) = e^{-z/c}/c\, dz$ on $[0,\infty)$. Although the constraint is critical, $\widehat{P}$ has no singular component.

(b) Consider almost the same problem where $e^{-z}\, dz$ is replaced by $R(dz) = \frac{a}{1+z^3} e^{-z}\, dz$ with a the unit mass normalizing constant and where the constraint (7.15) holds for some $c \geqslant c_o := \int_{[0,\infty)} \theta\, dR$. It is shown in [11], by means of arguments which are specific to the relative entropy, that (7.12) is not attained whenever c is large enough. Let us treat this example with the results of the present paper in hand.
One sees that the constraint is critical since $\int_{[0,\infty)} e^{yz} R(dz) < \infty$ if $y \leqslant 1$ and $+\infty$ if $y > 1$. Keeping the same notation as in (a), one has to solve (7.16). As $\int_{[0,\infty)} \theta\, dP_y$ is an increasing function of y, one obtains $\int_{[0,\infty)} \theta\, dP_y \leqslant \int_{[0,\infty)} \theta\, dP_1 = \int_{[0,\infty)} \frac{z}{1+z^3}\, dz / \int_{[0,\infty)} \frac{dz}{1+z^3} := c_* < \infty$ for all $y \leqslant 1$ and (7.16) has no solution for $c > c_*$. Nevertheless, the dual equality states that the value of the minimizing problem with $c \geqslant c_*$ is equal to $\sup_{y,s}\{yc + s - \int_{[0,\infty)} (e^{yz+s} - 1)\, R(dz)\} = \sup_y\{yc - \Lambda(y)\} = c - \Lambda(1)$ which is finite for all c. Therefore, the problem (7.14) is attained and Theorem 4.2 tells us that its solutions must have a nonzero singular part whenever $c > c_*$. More precisely, in this case the solutions $\hat{\ell}$ have the following form

$$\hat{\ell} = P_1 + \hat{\ell}^s$$

with $\hat{\ell}^s \in L_\tau^s$ and $\langle \hat{\ell}^s, \theta \rangle = c - c_*$. Keeping the notation of Example 4.7, it has been seen at (4.8) that $\int_{S_\epsilon} e^{yz} R(dz) = \infty$, for any $y > 1$ and $\epsilon > 0$. With $R(S_\epsilon) < \epsilon$, this shows that one can choose $S_\epsilon = [z_\epsilon, \infty)$ with $\lim_{\epsilon \to 0} z_\epsilon = +\infty$. With a lot of abuse, one might say that the "support" of $\hat{\ell}^s$ is "at infinity."
Note that $\hat{\ell}$ has a unit mass since $\langle \hat{\ell}^s, 1 \rangle = 0$. It also follows from these considerations that the corresponding problem (7.12) with the usual relative entropy shares the same unique solution with (7.14) when $c \leqslant c_*$ and has no solution when $c > c_*$.

(c) This is a variant of example (b). Consider du to be the uniform probability measure on $[0,1]$ and the minimization problem of the extended relative entropy $\bar{I}(\pi|du)$, $\pi \in \mathcal{E}([0,1])$ under the constraint $\int_{[0,1]} \Phi(u)\, \pi(du) \geqslant c$. Taking

$$\Phi = F^{-1} \tag{7.17}$$

to be the reciprocal of the distribution function $F(x) = \int_0^x \frac{a}{1+z^2} e^{-z}\, dz$ of R as in (b), one sees that the solution $\hat{\pi} = \hat{\ell} \# F$ admits a nonzero singular part if $c > c_*$ ($\hat{\ell} \# F$ is the image of $\hat{\ell}$ by the mapping F) and that the "support" of $\hat{\ell}^s$ is $\{1\}$ in the sense that one can choose $S_\epsilon = [1-\epsilon, 1]$.
By means of this example, one can show that for any probability measure R with an infinite support (so that $E_\tau(R) \subsetneqq L_\tau(R)$), one can find a real-valued constraint function θ which is critical and such that for some constraint interval $C = [c, \infty)$, (7.12) has no solution while (7.14) admits solutions with a nonzero singular component.
Clearly, this also holds for any γ^* such that λ does not satisfy the Δ_2-condition, see (2.6) and the comment below.

(d) This is a variant of example (c). Take the framework of example (b) at Section 7.3 with R the law of a Brownian motion on the unit circle $Z = \mathbb{R}/\mathbb{Z}$ with the uniform distribution du as its initial law. Clearly, this law is stationary: $R_t(du) = du$ for all $0 \leqslant t \leqslant 1$, where $R_t = R \# \zeta_t$ is the law of the position ζ_t at time t under R. In particular, $R_1(du) = du$. We look at problem (7.14) under the constraint

$$\langle \Phi(\zeta_1), \ell \rangle \geqslant c, \quad \ell \in \mathcal{E}(\mathcal{Z}).$$

As a consequence of (c) above, if Φ is given at (7.17) this constraint is critical.
On the other hand, if Φ satisfies $\int_{[0,1)} e^{\alpha|\Phi(u)|}\, du < \infty$ for all α, then the constraint is good and the unique solution to (7.12) and (7.14) is

$$\widehat{P}(\cdot) = \int_{\mathbb{R}/\mathbb{Z}} R^u(\cdot)\, \hat{\pi}(du) \tag{7.18}$$

where $R^u(dz)$ is the conditional law $R(dz|z_1 = u)$ (i.e. R uniquely disintegrates as $R(\cdot) = \int_{\mathbb{R}/\mathbb{Z}} R^u(\cdot)\, R_1(du)$) and $\hat{\pi}$ is the unique solution to

$$\textit{minimize } I(\pi|du) \quad \textit{subject to} \quad \int_{\mathbb{R}/\mathbb{Z}} \Phi\, d\pi \geqslant c, \quad \pi \in P_{\mathbb{R}/\mathbb{Z}},$$

whose solution is $\hat{\pi}(du) = \frac{1}{\mathcal{N}} e^{\tilde{\omega}\Phi(u)}\, du$ for some real $\tilde{\omega}$. The representation (7.18) is a direct consequence of the tensorization property of the relative entropy: $I(P|R) = I(P \# \zeta_1 | R \# \zeta_1) + \int_{\mathbb{R}/\mathbb{Z}} I(P^u|R^u) P \# \zeta_1(du)$ together with the fact that $I(P^u|R^u) = 0$ if and only if $P^u = R^u$.

References

[1] J.M. Borwein, On the failure of maximum entropy reconstruction for Fredholm equations and other infinite systems, Math. Program. 61 (1993) 251–261.
[2] J.M. Borwein, A.S. Lewis, Convergence of best entropy estimates, SIAM J. Optim. 1 (1991) 191–205.
[3] J.M. Borwein, A.S. Lewis, Duality relationships for entropy-like minimization problems, SIAM J. Control Optim. 29 (1991) 325–338.
[4] J.M. Borwein, A.S. Lewis, On the convergence of moment problems, Trans. Amer. Math. Soc. 325 (1991) 249–271.
[5] J.M. Borwein, A.S. Lewis, Partially-finite programming in l_1 and the existence of the maximum entropy estimates, SIAM J. Optim. 3 (1993) 248–267.
[6] J.M. Borwein, A.S. Lewis, Strong rotundity and optimization, SIAM J. Optim. 1 (1994) 146–158.
[7] J.M. Borwein, A.S. Lewis, R.D. Nussbaum, Entropy minimization, DAD problems and doubly stochastic kernels, J. Funct. Anal. 123 (1994) 264–307.
[8] N. Bourbaki, Eléments de mathématique. Intégration 1–4, second ed., Hermann, Paris, 1961.
[9] A.B. Cruzeiro, L. Wu, J.C. Zambrini, Bernstein processes associated with a Markov process, in: R. Rebolledo (Ed.), Stochastic Analysis and Mathematical Physics, ANESTOC'98. Proceedings of the Third International Workshop, in: Trends Math., Birkhäuser, Boston, 2000, pp. 41–71.
[10] I. Csiszár, I-divergence geometry of probability distributions and minimization problems, Ann. Probab. 3 (1975) 146–158.
[11] I. Csiszár, Sanov property, generalized I-projection and a conditional limit theorem, Ann. Probab. 12 (1984) 768–793.
[12] I. Csiszár, F. Gamboa, E. Gassiat, MEM pixel correlated solutions for generalized moment and interpolation problems, IEEE Trans. Inform. Theory 45 (7) (1999) 2253–2270.
[13] D. Dacunha-Castelle, F. Gamboa, Maximum d'entropie et problème des moments, Ann. Inst. H. Poincaré Probab. Statist. 26 (1990) 567–596.
[14] H. Föllmer, Random Fields and Diffusion Processes, Ecole d'Eté de Probabilités de Saint-Flour XV-XVII-1985-87, Lecture Notes in Math., vol. 1362, Springer, Berlin, 1988.
[15] F. Gamboa, E. Gassiat, Bayesian methods and maximum entropy for ill-posed inverse problems, Ann. Statist. 25 (1) (1997) 328–350.
[16] H. Gzyl, The Method of Maximum Entropy, World Scientific, 1994.
[17] E.T. Jaynes, Papers on Probability, Statistics and Statistical Physics, Reidel, Dordrecht, 1983.
[18] A. Kozek, Orlicz spaces of functions with values in Banach spaces, Ann. Soc. Math. Polonae, Ser. I, Comment. Math. XIX (1977) 259–288.
[19] A. Kozek, Convex integral functionals on Orlicz spaces, Ann. Soc. Math. Polonae, Ser. I, Comment. Math. XXI (1979) 109–135.
[20] C. Léonard, Convex minimization problems with weak constraint qualifications, preprint.
[21] C. Léonard, Entropic projections and dominating points, preprint.
[22] C. Léonard, Convex conjugates of integral functionals, Acta Math. Hungar. 93 (4) (2001) 253–280.
[23] C. Léonard, Minimizers of energy functionals, Acta Math. Hungar. 93 (4) (2001) 281–325.
[24] C. Léonard, Minimizers of energy functionals under not very integrable constraints, J. Convex Anal. 10 (1) (2003) 63–88.
[25] M.M. Rao, Z.D. Ren, Theory of Orlicz Spaces, Pure Appl. Math., vol. 146, Marcel Dekker, Inc., 1991.
[26] R.T. Rockafellar, Convex integral functionals and duality, in: E. Zarantonello (Ed.), Contributions to Nonlinear Functional Analysis, Academic Press, New York, 1971, pp. 215–235.
[27] R.T. Rockafellar, Conjugate Duality and Optimization, Regional Conf. Ser. in Appl. Math., vol. 16, SIAM, Philadelphia, 1974.
[28] E. Schrödinger, Sur la théorie relativiste de l'électron et l'interprétation de la mécanique quantique, Ann. Inst. H. Poincaré 2 (1932) 269–310.
[29] C.R. Smith, G.J. Erickson, P.O. Neudorfer (Eds.), Maximum Entropy and Bayesian Methods, Proceedings of the 11th International Workshop on Maximum Entropy and Bayesian Methods of Statistical Analysis, Kluwer, Seattle, 1991.
[30] M. Teboulle, I. Vajda, Convergence of best ϕ-entropy estimates, IEEE Trans. Inform. Theory 39 (1993) 297–301.