ECOLE POLYTECHNIQUE

CENTRE DE MATHÉMATIQUES APPLIQUÉES
UMR CNRS 7641

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

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

A Convex Optimization Problem Arising From Probabilistic Questions.

Christian Léonard

R.I. N^0 506 *May 2003*

A CONVEX OPTIMIZATION PROBLEM ARISING FROM PROBABILISTIC QUESTIONS

CHRISTIAN LÉONARD

Abstract. An abstract convex minimization problem is solved by means of the classical conjugate duality. No topological constraint qualifications are imposed and the representation of the minimizers is obtained in great generality. Such a minimization naturally arises in minimum entropy methods, mass transportation problems and in large deviations theory when looking at conditional laws of large numbers. Some applications are derived in the area of variational processes and Markov processes.

1. Introduction

1.1. **A few motivations.** Let $\{L_n\}$ be a collection of random vectors which obeys a Large Deviation Principle (LDP) in some topological space $\mathcal{L}$ with the good rate function I. Suppose that I is strictly convex and let Q_o be the unique global minimizer of I : $I(Q_o) = 0$. Then, the sequence (L_n) tends in law (and almost surely in some probability space) to Q_o. Let us consider a sequence of conditioning events $\{L_n \in A_0\}$ for some measurable convex set A_0 and suppose that $\mathbb{P}(L_n \in A_0)$ is positive for all $n \geq 1$ (this will be relaxed later). The Conditional Law of Large Numbers states that conditionally on $L_n \in A_0$ for all n, the sequence (L_n) tends in law to the deterministic vector $\bar{Q}$ which is the unique minimizer of $I(Q)$ subject to the constraint $Q \in A_0$. More precisely, this means that $\mathbb{P}(L_n \in \cdot \mid L_n \in A_0)$ converges weakly to the Dirac measure at $\bar{Q} : \delta_{\bar{Q}}$. This simple fact is the foundation of many important convergence results in statistical physics such as the Gibbs Conditioning Principle. The analytic counterpart of this probabilistic question is the optimization problem

$$\text{minimize } I(Q) \text{ subject to } Q \in A_0, Q \in \mathcal{L}. \tag{1}$$

On the other hand, suppose for instance that $I(Q)$ is the relative entropy $I(Q \mid R)$ of the probability measure Q with respect to some reference probability measure R. If some criterion insures the existence of a solution to (1), it implies a fortiori that there exists at least a probability measure in A_0 which is absolutely continuous with respect to R. Typical examples of constraints $Q \in A_0$ are *moment constraints:* $\int f_i \, dQ \in C_i$, $i \in \mathcal{I}$ or *marginal constraints:* $Q_a \in C_a$ and $Q_b \in C_b$ where Q_a and Q_b are the marginal measures of Q on the product space $\Omega_a \times \Omega_b$.

Another interesting problem is the Monge-Kantorovitch mass transportation problem. It corresponds to $I(Q) = \int_{\Omega_a \times \Omega_b} c(\omega_a, \omega_b) \, Q(d\omega_a d\omega_b)$ where c is interpreted as a cost function, Q is a probability measure on the product space $\Omega_a \times \Omega_b$ and $Q \in A_0$ is given by the above marginal constraints.

Date: May 2003.

1991 *Mathematics Subject Classification.* Primary: , Secondary:

Key words and phrases. Large deviations.

In the present paper, the above convex minimization problem is studied in great details. Our approach is based on the very well known theory of conjugate duality applied to convex optimization as developped in Rockafellar's monograph [27]. Our abstract results are also applied to specific probabilistic questions in relation to *minimum entropy* methods, *conditional laws of large numbers* and the construction of *constrained variational processes* such as Bernstein's and Nelson's processes.

1.2. **Applying conjugate duality.** Conjugate duality is a powerful tool to solve problems of the type of (1), provided that I and A_0 are convex. Assume that $I=\Gamma^*$ is the convex conjugate of some convex function $\Gamma:\mathcal{U}\to(-\infty,\infty]$ on a vector space $\mathcal{U}$ in duality with $\mathcal{L}$. Also assume that $Q\in A_0$ means that $TQ\in C_0$ where $T:\mathcal{L}\to\mathcal{X}$ is a linear operator with values in another vector space $\mathcal{X}$. Then (1) becomes

$$\textit{minimize } \Gamma^*(Q) \textit{ subject to } TQ\in C_0, Q\in\mathcal{L}. \tag{2}$$

Let $\mathcal{Y}$ be a vector space in duality with $\mathcal{X}$ and define $T^T:\mathcal{Y}\to\mathcal{U}$ as the operator adjoint to T. The dual problem of (2) is the unconstrained maximization problem:

$$\textit{maximize } \inf_{x\in C_0}\langle x,y\rangle-\Gamma(T^Ty), y\in\mathcal{Y} \tag{3}$$

Let us illustrate this conjugate duality by means of a classical example in large deviation theory. Let $L_n=\frac{1}{n}\sum_{i=1}^n\delta_{Z_i}\in\mathcal{L}$ be the empirical measure of an R-iid sample $(Z_i)_{i\geq1}$ with values in a vector space $\mathcal{X}$ and let $X_n=\frac{1}{n}\sum_{i=1}^n Z_i\in\mathcal{X}$ be its empirical mean. By Sanov's theorem, $\{L_n\}$ obeys the LDP with the relative entropy $I(Q)=I(Q\mid R)$ as rate function, while by Cramér's theorem $\{X_n\}$ obeys the LDP with rate function $J(x)=\sup_{y\in\mathcal{Y}}\{\langle x,y\rangle-\log\int e^{\langle y,x\rangle}\,R(dx)\}$, $x\in\mathcal{X}$. Not getting into the details, think of $\mathcal{L}$ as the dual space of some space $\mathcal{U}$ of measurable functions on $\mathcal{X}$. Consider the operator $TQ=\int x\,Q(dx)$, so that $TL_n=X_n$ and $T^Ty\in\mathcal{U}$ is defined by $T^Ty(x)=\langle y,x\rangle$, $x\in\mathcal{X}$. It appears that taking $\Gamma(u)=\log\int e^{u(x)}\,R(dx)$ for $u\in\mathcal{U}$, we have $I=\Gamma^*$ for the duality $(\mathcal{L},\mathcal{U})$. Taking $C_0=\{x_o\}$ in the dual problem, we also obtain $\sup(3)=J(x_o)$. More, by the contraction principle, we get $J(x_o)=\inf\{I(Q);Q,TQ=x_o\}$. This equality is $\inf(2)=\sup(3)$ with $C_0=\{x_o\}$.
With C_0 a general subset of $\mathcal{X}$, the solution to (2) is called the *generalized I-projection* of R onto $A_0=\{Q;TQ\in C_0\}$ and the solution to

$$\textit{minimize } J(x) \textit{ subject to } x\in C_0, x\in\mathcal{X} \tag{4}$$

is called the *dominationg point* of C_0 with respect to the rate function J.

In regular situations, we obtain the equality of the values of these problems: $\inf(2)=\sup(3)$. In particular, $\sup(3)<\infty$ is a variational criterion for the existence of a solution to (2). If the constraint operator T has a finite dimensional range $\mathcal{X}$, then (3) is a finite dimensional unconstrained problem whose solution $\bar{y}$ is an inward normal vector to C_0 and is linked to the solution $\bar{Q}$ to (1) by the formal identity:

$$\bar{Q}=\Gamma'(T^T\bar{y}). \tag{5}$$

If the constraint is infinite dimensional, several spaces $\mathcal{Y}$ may be dually linked to $\mathcal{X}$, so that several dual problems may be considered. In order that a representation formula (5) holds, one has to consider the largest possible $\mathcal{Y}$: the algebraic dual space $\mathcal{X}^\sharp$ of $\mathcal{X}$. This shifts the difficulty to the computation of the convex conjugate Γ of $I=\Gamma^*$ for the saturated duality $(\mathcal{X}, \mathcal{X}^\sharp)$. We are going to implement this program in the case where I is an *entropy*, that is a convex integral functional of the form

$$I(Q) = \int_\Omega \rho^*(\frac{dQ}{dR})\, dR \tag{6}$$

with ρ^* a convex function and R a reference probability measure on some measure space Ω.

In this paper, based on the theory of conjugate duality, we solve the primal problem (2) via the dual problem (3). Although this approach is highly classical, our implementation is unusual since we consider the saturated duality $(\mathcal{X}, \mathcal{Y} = \mathcal{X}^\sharp)$ to build the dual problem. This allows us not to assume any a priori topological regularity on the constraint (no topological qualification of the constraints).

1.3. **Outline of the paper.** In Section 2, an abstract conditional law of large numbers is stated to motivate the minimization problem (1). Some details are also given about *randomly weighted means* whose large deviations admit entropies of the type of (6) as rate functions.

Section 3 is devoted to the statements of the main results about the convex minimization problem (2). These results are Theorem 3.2: dual equalities and primal attainment, Theorem 3.3: dual attainment, and Theorem 3.4: dual representation of the minimizers.

These abstract results are applied in Section 4 to the minimization of entropies of the type of (6) in the context of stochastic processes with marginal constraints. In particular we look at the case where the initial and final laws of the processes are prescribed. These variational processes are called Bernstein processes. They are the foundation of the Euclidean Quantum Mechanics initialized by Schrödinger [29] and developped by Zambrini [33]. We also consider the situation where a whole flow of t-marginals is prescribed. This gives rise to another kind of variational processes called Nelson processes. They are the foundation of the Stochastic Mechanics initialized by E. Nelson. At the end of this section, at Theorem 4.2, it is proved that minimizing the relative entropy $I(Q \mid R)$ under some well-suited constraints preserves the conditional independence properties of the reference measure R. For instance, when minimizing the relative entropy with respect to some Markov law R to build Bernstein's or Nelson's processes, one obtains a Markov process (which is absolutely continuous w.r.t. R) as a solution.

Section 5 is a short review of the main results of conjugate duality. It is there for the convenience of the reader and allow us to give self-contained proofs of the results of Section 3 at Section 6. Finally, the easy proof of Theorem 2.1 is given at Section 7.

1.4. **About the literature.** A well-known application of conditional laws of large numbers is the Gibbs Conditioning Principle. For a clear account on this subject, one may read ([13], Section 7.3).

The literature about the minimization of entropy functionals under convex constraints is considerable: many papers are concerned with an engineering approach, working on the implementation of numerical procedures in specific situations. In fact, entropy minimization is a popular method to solve ill-posed inverse problems. For connections with statistics, see for instance [22].

Surprisingly enough, rigorous general results on this topic are quite recent. Let us cite,

among others, the main contribution of Borwein and Lewis: [1], [2], [3], [4], [5], [6] together with the paper [32] by Teboulle and Vajda. In these papers, topological constraint qualifications are assumed: it is assumed that the constraints stand in some topological interior of the domain of I. Such restrictions are removed in the present article.
In the special case where I is the relative entropy, Csiszár has obtained the best existence results in [9] together with powerful dual equalities. His proofs are based on geometrical properties of the relative entropy; no convex analysis is needed. Based on the same geometrical ideas, the same author has obtained later in [10] a powerful Gibbs Conditioning Principle for non-interacting particles.

Although we provide a new proof for the representation of Bernstein's and Nelson's processes, the entropic approach to build variational processes is already known. This technique is developped by Cattiaux and Léonard in [8], in connection with Sanov's and Cramér's theorems, to build Nelson's processes. The same approach is used by Cruzeiro, Wu and Zambrini in [12] to build Bernstein's processes. When displaying at Section 4 these applications, our aim was to give a non trivial illustration in stochastic analysis of the general results of Section 3. On the other hand, Theorem 4.2 is a new result.

As already alluded to, another natural probabilistic application of the abstract convex minimization problems (2) and (4) is the theory of generalized I-projections and dominating points. Generalized projections are studied by Csiszár in [10] and [11]. Dominating points have been introduced by Ney in [23] and [24]. Their properties are also investigated by Einmahl and Kuelbs in [14] and Kuelbs in [17]. We extend and explain these results in the light of the results of Section 3 in [20].

Mass transportation problems are extensively investigated in the monograph [25] by Rachev and Rüschendorf. Let us mention that the proof of the celebrated Strassen's theorem on the existence of probabilities with given marginals [30] is based on the conjugate duality of some minimization problem (2) related to a mass transportation problem.

1.5. **Some notations.** ?

2. Probabilistic questions

In this section, we state at Theorem 2.1 conditional laws of large numbers. It is an easy result, which motivates the study of a naturally associated convex minimization problem, see (7). This minimization problem will be solved in great generality at Section 3.

2.1. **Conditional laws of large numbers.**

Let $\{L_n\}$ be a sequence of random vectors in the algebraic dual space $\mathcal{L}_0$ of some vector space $\mathcal{U}_0$. As a typical instance, one can think of random measures L_n with $\mathcal{U}_0$ a function space. Let $T : \mathcal{L}_0 \to \mathcal{X}_0$ be a linear operator with values in another vector space $\mathcal{X}_0$. We are going to investigate the behavior of the conditional law $I\!P(L_n \in \cdot \mid TL_n \in C_0)$ of L_n as n tends to infinity, for some measurable set C_0 in $\mathcal{X}_0$. It appears that this type of conditional law of large numbers is connected with large deviations. We assume that $\{L_n\}$ obeys the LDP with a good rate function I in $\mathcal{L}_0$ endowed with the weak topology $\sigma(\mathcal{L}_0, \mathcal{U}_0)$ and the associated Borel σ-field. It is also clear that one should assume that $I\!P(TL_n \in C_0) > 0$ for all n, not to divide by zero. To overcome this restriction, we look at $I\!P(L_n \in \cdot \mid TL_n \in C_\delta)$ where C_δ tends to C_0 as δ tends to zero.

Let us assume that $\mathcal{X}_0$ is a topological vector space with its Borel σ-field and that $T: \mathcal{L}_0 \to \mathcal{X}_0$ is continuous. The contraction principle tells us that

$$X_n \triangleq TL_n$$

obeys the LDP in $\mathcal{X}_0$ with the rate function

$$J(x) = \inf\{I(\ell); \ell \in \mathcal{L}_0, T\ell = x\}.$$

If $L_n = \frac{1}{n}\sum_{i=1}^n \delta_{Z_i}$ is the empirical measure of an iid sequence (Z_i) of $\mathcal{X}_0$-valued random variables and $X_n = \frac{1}{n}\sum_{i=1}^n Z_i$ is its empirical mean, the large deviations for L_n (taking for $\mathcal{U}_0$ the space of bounded measurable functions) and X_n are described by *Sanov*'s and *Cramér*'s theorems.

We make the following

Assumptions on $\{L_n\}$. *The sequence* $\{L_n\}$ *obeys the LDP in* $\mathcal{L}_0$ *with a* good *rate function* I. *This means that* I *is inf-compact.*

As a convention, one writes $J(C)$ for $\inf_{x\in C} J(x)$.
Assumptions on the conditioning event.

(a) *The linear operator* $T : \mathcal{L}_0 \to \mathcal{X}_0$ *is continuous.*
(b) $J(C_0) < \infty$.
(c) *The set* C_0 *is closed, it is the limit as* δ *decreases to 0:* $C_0 \triangleq \cap_\delta \mathrm{cl}\ C_\delta$, *of the closures of a nonincreasing family of Borel sets* C_δ *in* $\mathcal{X}_0$ *such that for all* $\delta > 0$ *and all* $n \geq 1$, $\mathbb{P}(X_n \in C_\delta) > 0$
(d) *and one of the following statements*
  1. $C_\delta = C_0$ *for all* $\delta > 0$ *and* $J(\mathrm{int}\ C_0) = J(C_0)$, *or*
  2. $C_0 \subset \mathrm{int}\ C_\delta$ *for all* $\delta > 0$.

  is fulfilled.

This framework is based on [31] and ([13], Section 7.3).

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

$$\textit{minimize } I(\ell) \textit{ subject to } T\ell \in C_0, \quad \ell \in \mathcal{L}_0. \tag{7}$$

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

$$\textit{minimize } J(x) \textit{ subject to } x \in C_0, \quad x \in \mathcal{X}_0. \tag{8}$$

We can now state a result about conditional laws of large numbers. The easy proof of this theorem is postponed to Section 7.

Theorem 2.1. *For all open subset* G *of* $\mathcal{L}_0$ *such that* $\mathcal{G} \subset G$ *and all open subset* H *of* $\mathcal{X}_0$ *such that* $\mathcal{H} \subset H$, *we have*

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

In particular, if C_0 is convex and the rate functions I and J are strictly convex, we have the conditional laws of large numbers:

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

where the limits are understood with respect to the usual weak topologies of probability measures and $\bar\ell$ is the unique solution to the convex minimization problem (7) and $\bar x = T\bar\ell$ is the unique solution to (8).

In the rest of this paper, one will only work with functionals I which are the convex conjugates of some function Φ on $\mathcal{U}_0$. This means that

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

In the following examples, $\mathcal{U}_0$ is a space of measurable functions $u : \Omega \to \mathbb{R}$ on a measurable space Ω. Consequently $\mathcal{L}_0$ may contain measures on Ω.

Empirical means of an iid sequence. In the special situation where

$$L_n = \frac{1}{n}\sum_{i=1}^n \delta_{Z_i}$$

is the empirical measure of an iid sequence (Z_i) of $\mathcal{X}_0$-valued random variables and $X_n = \frac{1}{n}\sum_{i=1}^n Z_i$ is its empirical mean, I is the *relative entropy* with respect to $R \triangleq \mathcal{L}aw(Z)$: the common law of the Z_i,'s and J is the convex conjugate of the log-Laplace transform of R. We have: $I = \Phi^*$ with $\Phi(u) = \log \mathbb{E}e^{u(Z)}$, $u \in \mathcal{U}_0$ and $J = \Lambda^*$ with $\Lambda(y) = \log \mathbb{E}e^{\langle y,Z\rangle}$, $y \in \mathcal{Y}_0$ where $\mathcal{Y}_0$ is a vector space in duality with $\mathcal{X}_0$.

The notation R is chosen to remind that R is a *reference measure.*

The vector $\bar x$ is called a *dominating point* of C_0 for the empirical mean of the Z_i's and the measure $\bar\ell$ is the *I-projection* (in the sense of Csiszár) of R on the convex set $\{\ell$ probability measure on $\mathcal{X}_0; \int_{\mathcal{X}_0} z\,\ell(dz) \in C_0\}$.

Randomly weighted means. Let us consider a deterministic triangular array $\{(\omega_i^n)_{1\le i\le n}, n \ge 1\}$ of elements ω_i^n in a measurable space Ω such that the empirical measure $R_n \triangleq \frac{1}{n}\sum_{i=1}^n \delta_{\omega_i^n}$ tends to some probability measure R as n tends to infinity (in some weak sense). A random weight W_ω is attached to each $\omega \in \Omega$: $\{W_\omega \in \Omega\}$ is a family of independent real-valued random variables. Under some assumptions, the large deviations as n tends to infinity of the sequence of random signed measures on Ω :

$$L_n \triangleq \frac{1}{n}\sum_{i=1}^n W_{\omega_i^n}\delta_{\omega_i^n}, n \ge 1$$

is governed by the rate function

$$I(\ell) = \int_\Omega \gamma^*(\omega, \frac{d\ell}{dR}(\omega))\, R(d\omega) \tag{9}$$

if $\ell \in \mathcal{L}_0$ is an absolutely continuous measure with respect to R and $I(\ell) = +\infty$ otherwise, where $\gamma^*(\omega, t)$, $t \in \mathbb{R}$ is the convex conjugate of $\gamma(\omega, s) = \log \mathbb{E}e^{sW_\omega}$, $s \in \mathbb{R}$.

The large deviations of randomly weighted means have been studied in the context of statistical physics in [15] and in statistics in [16]. For a proof of the LDP one may read

([13], Theorem 7.2.3).
When γ doesn't depend on ω, the functional I is sometimes called *γ^*-entropy.* An important instance is the *relative entropy* which corresponds to $\gamma(s) = e^s - 1$: the log-Laplace transform of the Poisson law with unit expectation. Indeed, $\gamma^*(t) = t\log t - t + 1$ if $t \geq 0$ and $\gamma^*(t) = +\infty$ if $t < 0$. Therefore, requiring the constraint $\langle \ell, \mathbf{1} \rangle = 1$, one obtains $I(\ell) = \int_\Omega \log(\frac{d\ell}{dR})\, dR$ if ℓ is a probability measure which is absolutely continuous with respect to R and $+\infty$ otherwise. In other words, under this unit mass constraint $I(\ell)$ is the relative entropy of ℓ with respect to the reference measure R.
It appears that (provided that $\mathcal{U}_0$ is large enough) $I = \Phi^*$ is the convex conjugate of $\Phi(u) = \int_\Omega \gamma(\omega, u(\omega))\, R(d\omega)$, $u \in \mathcal{U}_0$.

2.2. **Some constraints.** The conditioning events $TL_n \in C_\delta$ become $T\ell \in C_0$ in the minimization problem (7) as n tends to infinity and δ tends to zero. Now, one can interpret $T\ell \in C_0$ as a constraint for the minimization problem.
As before, $\mathcal{U}_0$ is a space of measurable functions $u : \Omega \to I\!R$ on a measurable space Ω and $\mathcal{L}_0$ may contain measures on Ω.
Moment constraints. Let $\theta_1, \dots, \theta_n$ be n functions on Ω which belong to the space $\mathcal{U}_0$. The constraint $\langle \ell, \theta_i \rangle \in A_i$ for all $1 \leq i \leq n$ where the A_i's are intervals is written: $T\ell \in C_0$ with $T\ell = (\langle \ell, \theta_i \rangle)_{1\leq i\leq n} \in \mathcal{X}_0 = I\!R^n$ and $C_0 = A_1 \times \cdots \times A_n \subset I\!R^n$.

The problems of minimization of γ^*-entropies under moment constraints are widely investigated in numerous domains of applied mathematics, including statistical physics and ill-posed inverse problems, see [22] for instance.

Marginal constraints. Let $\Omega = \Omega_0 \times \Omega_1$ be a product space. If ℓ is a probability measure on $\Omega_0 \times \Omega_1$, its marginals ℓ_0 on Ω_0 and ℓ_1 on Ω_1 are defined by $\langle \ell_0, u_0 \rangle = \langle \ell, u_0 \otimes \mathbf{1} \rangle$ and $\langle \ell_1, u_1 \rangle = \langle \ell, \mathbf{1} \otimes u_1 \rangle$ for all bounded measurable functions u_0 on Ω_0 and u_1 on Ω_1. Therefore the marginal couple $(\ell_0, \ell_1) = T\ell$ is specified by infinitely many moment constraints.

Minimizing a γ^*-entropy under marginal constraint is a well known problem. A large deviation approach to this problem is used by Cattiaux and Gamboa in [7] and a complete answer to this problem with a convex analytical approach which differs from this paper's one, is given by the author in [19].

2.3. **The Monge-Kantorovich problem.** This problem enters a class of problems which isn't directly connected to Large Deviations. It is a relaxed form of the original Monge transportation problem. It consits in minimizing the functional $\ell \mapsto \int_{\Omega_0\times\Omega_1} c(\omega_0, \omega_1)\, \ell(d\omega_0 d\omega_1)$ where $c : \Omega_0 \times \Omega_1 \to [0, +\infty]$ is a cost function, among all the probability measures ℓ on $\Omega_0 \times \Omega_1$ with prescribed marginals (ℓ_0, ℓ_1). This problem enters our framework taking $\Phi(u) = 0$ if $u \leq c$ and $+\infty$ otherwise. Many recent papers are concerned with this problem. For a comprehensive account on this subject see [25].

3. An abstract convex minimization problem

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

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

We shall be concerned with the following convex minimization problem

$$\textit{minimize } \Phi^*(\ell) \textit{ subject to } T\ell \in C_0,\ \ell \in \mathcal{L}_0 \tag{P_0}$$

where C_0 is a convex subset of a vector space $\mathcal{X}_0$ and $T : \mathcal{L}_0 \to \mathcal{X}_0$ is a linear operator.

The problem (7) is (P_0) in the situation where $I = \Phi^*$ is a closed convex function, which is a common feature when independent variables are considered, and C_0 is a convex set.

3.1. **The operator T and its adjoint.** Let $\mathcal{X}_0$ be in separating duality with a vector space $\mathcal{Y}_0$. It is often useful to define the constraint operator T by means of its adjoint $T^T : \mathcal{Y}_0 \to \mathcal{L}_0^\sharp$ ($\mathcal{L}_0^\sharp$ is the algebraic dual space of $\mathcal{L}_0$), as follows. For all $\ell \in \mathcal{L}_0, x \in \mathcal{X}_0$,

$$T\ell = x \overset{\Delta}{\iff} \forall y \in \mathcal{Y}_0, \langle T^T y, \ell\rangle_{\mathcal{L}_0^\sharp,\mathcal{L}_0} = \langle x, y\rangle_{\mathcal{X}_0,\mathcal{Y}_0}.$$

We shall assume that the restriction $T^T(\mathcal{Y}_0) \subset \mathcal{U}_0$ holds, where $\mathcal{U}_0$ is identified with a subspace of $\mathcal{L}_0^\sharp = \mathcal{U}_0^{\sharp\sharp}$. This provides us with the diagram

$$\begin{array}{ccc} \Big\langle\ \mathcal{U}_0 & , & \mathcal{L}_0\ \Big\rangle \\ T^T\Big\uparrow & & \Big\downarrow T \\ \Big\langle\ \mathcal{Y}_0 & , & \mathcal{X}_0\ \Big\rangle \end{array} \tag{Schema 0}$$

For the moment constraints $T\ell = (\langle \ell, \theta_i\rangle)_{1\le i\le n} \in \mathcal{X}_0 = I\!R^n$ we have: $T^T y = \sum_{i=1}^n y_i\theta_i \overset{\Delta}{=} \langle y, \theta(\cdot)\rangle$, $y \in \mathcal{Y}_0 = I\!R^n$ with $\theta(\omega) = (\theta_i(\omega))_{1\le i\le n} \in \mathcal{X}_0$.

In the setting of the marginal constraints, the set $\mathcal{X}_0 = \mathcal{P}(\Omega_0) \times \mathcal{P}(\Omega_1)$ is dually linked with $\mathcal{Y}_0 = B(\Omega_0) \times B(\Omega_1)$ where $\mathcal{P}(\Omega)$ is set of all probability measures on Ω and $B(\Omega)$ is the space of all measurable bounded functions on Ω. The constraint operator is $T\ell = (\ell_0, \ell_1) \in \mathcal{X}_0$ and for any $y = (y_0, y_1) \in B(\Omega_0) \times B(\Omega_1)$ we have $[T^T y](\omega_0, \omega_1) = y_0(\omega_0) + y_1(\omega_1)$, that is $T^T y = y_0 \oplus y_1$. Let us notice that as for the moment constraints one can write $T^T y = \langle y, \theta(\cdot)\rangle$ with $\theta(\omega_0, \omega_1) = (\delta_{\omega_0}, \delta_{\omega_1}) \in \mathcal{X}_0$ since $\langle y, \theta(\cdot)\rangle = \langle (y_0, y_1), (\delta_{\omega_0}, \delta_{\omega_1})\rangle = \int_{\Omega_0} y_0\, \delta_{\omega_0} + \int_{\Omega_1} y_1\, \delta_{\omega_1} = y_0(\omega_0) + y_1(\omega_1)$.

3.2. **A formal Lagrangian approach.** The existence of a solution to (P_0) implies that its *value*

$$\inf(P_0) \overset{\Delta}{=} \inf\{\Phi^*(\ell); T\ell \in C_0, \ell \in \mathcal{L}_0\}$$

is finite: $\inf(P_0) < +\infty$. It is often useful to have an alternate variational expression for this value. Such an expression

$$\inf(P_0) = \sup(D_0)$$

is called a *dual equality* where (D_0) is a maximization problem dually linked to (P_0). The minimization problem (P_0) is called the *primal problem* and the maximization problem is called its *dual problem.* In some situations, $\sup(D_0) < +\infty$ is easier to obtain than $\inf(P_0) < +\infty$.

In the following we are going to introduce formally the problem (D_0). The Lagrange multipliers method mainly states that as $A_0 \overset{\Delta}{=} \{\ell \in \mathcal{L}_0; T\ell \in C_0\}$ is a convex subset of $\mathcal{L}_0$ since C_0 is a convex subset of $\mathcal{X}_0$, $\bar\ell$ minimizes the convex function Φ^* on A_0 if the derivative $\Phi^{*\prime}(\bar\ell)$ is an inward normal of A_0 at $\bar\ell$. Hence, let us take our Lagrangian of the following form: $\Phi^*(\ell) - [\langle \ell, u\rangle - \alpha(u)]$ where α is a concave function. If $(\bar\ell, \bar u)$ is a saddle-point, cancelling the partial derivatives of the Lagrangian we obtain $\Phi^{*\prime}(\bar\ell) = \bar u$ and $\bar\ell \in \partial\alpha(\bar u)$ (the superdifferential of α at $\bar u$.) Therefore we want $\bar u$ to be an inward normal of A_0 at $\bar\ell$ and $\partial\alpha(\bar u) \subset A_0$. One can choose for α the function $\alpha(u) = \inf_{\ell\in A_0}\langle \ell, u\rangle$. But, $\inf_{\ell\in A_0}\langle \ell, u\rangle = \inf_{x\in C_0}\inf\{\langle \ell, u\rangle; T\ell = x, \ell \in \mathcal{L}_0\}$ and in order that $\inf\{\langle \ell, u\rangle; T\ell = x, \ell \in \mathcal{L}_0\}$ is finite, it is necessary that for all ℓ, $T\ell = 0$ implies that $\langle \ell, u\rangle = 0$. Formally, this is obtained with $u = T^T y$, for some y. Finally, we have got $\alpha(u) = \alpha(T^T y) = \inf_{x\in C_0}\langle x, y\rangle$ and the natural Lagrangian for the problem (P_0) is $\Phi^*(\ell) - \langle T\ell, y\rangle + \inf_{x\in C_0}\langle x, y\rangle$, $\ell \in \mathcal{L}_0$, $y \in \mathcal{Y}_0$. The identity $\Phi^{*\prime}(\bar\ell) = \bar u$ becomes $\Phi^{*\prime}(\bar\ell) = T^T\bar y$ and the corresponding Young equality is $\Phi^*(\bar\ell) - \langle T\bar\ell, \bar y\rangle = -\Phi^{**}(T^T\bar y)$ where Φ^{**} is the convex biconjugate of Φ for the duality $\langle \mathcal{U}_0, \mathcal{L}_0\rangle$. Therefore, the natural dual problem associated with (P_0) is of the following form

$$\textit{maximize} \inf_{x\in C_0}\langle x, y\rangle - \Lambda(y), y \in \mathcal{Y}_0 \qquad (D_0)$$

where $\Lambda(y) \triangleq \Phi^{**}(T^T y)$. The desired variational formulation $\inf(P_0) = \sup(D_0)$ is of the form

$$\inf\{\Phi^*(\ell); T\ell \in C_0, \ell \in \mathcal{L}_0\} = \sup_{y\in\mathcal{Y}_0}\{\inf_{x\in C_0}\langle x, y\rangle - \Lambda(y)\}.$$

Moreover, the formal identity $\Phi^{*\prime}(\bar\ell) = T^T\bar y$ leads us via convex duality to the formal representation formula of the solutions of (P_0)

$$\bar\ell \in \partial\Phi(T^T\bar y).$$

3.3. **Main questions.** The main natural questions arising with (P_0) and (D_0) are related to

- the dual equality: Does $\inf(P_0) = \sup(D_0)$ hold?
- the primal attainment: Does there exist a solution $\bar\ell$ to (P_0)? What about the minimizing sequences, if any?
- the dual attainment: Does there exist a solution $\bar y$ to (D_0)?
- the representation of the primal solutions: Find an identity of the type of $\bar\ell \in \partial\Phi(T^T\bar y)$.

It appears that when the constraints are infinite-dimensional one can choose many different $\mathcal{Y}_0$'s without modifying (P_0). So that for a small set $\mathcal{Y}_0$ the dual attainment is not the rule. As a consequence, we are facing the problem of finding an extension of (D_0) such that the extended dual problem admits solutions in generic cases and the representation of the primal solution is $\bar\ell \in \partial\bar\Phi(T^T\bar y)$ where $\bar\Phi$ is an extension of Φ.

We are going to give an answer to these questions. Our strategy is to use convex duality as developed by Rockafellar in [27]. The main features of this approach are recalled in Section 5 for the reader's convenience.

We have taken as a rule not to introduce arbitrary topological assumptions since (P_0) is expressed without any topological notion. The convexity of the problem will allow us to take advantage of geometrical easy properties: the topologies to be considered later are related to gauges of some level sets of convex functions.

3.4. **Assumptions.** Let us give a list of our main hypotheses.

(H_Φ) 1- $\Phi : \mathcal{U}_0 \to [0, +\infty]$ is convex and $\Phi(0) = 0$
2- $\forall u \in \mathcal{U}_0, \exists \alpha > 0, \Phi(\alpha u) < \infty$
3- $\forall u \in \mathcal{U}_0, u \neq 0, \exists t \in I\!R, \Phi(tu) > 0$
(H_T) 1- $T^T(\mathcal{Y}_0) \subset \mathcal{U}_0$
2- $\ker T^T = \{0\}$
(H_C) $C \triangleq C_0 \cap \mathcal{X}$ is a convex $\sigma(\mathcal{X}, \mathcal{Y}_1)$-closed subset of $\mathcal{X}$

The definitions of the vector spaces $\mathcal{X}$ and $\mathcal{Y}_1$ which appear in the last assumption are stated below at Section 3.5. For the moment, let us only say that if C_0 is convex and $\sigma(\mathcal{X}_0, \mathcal{Y}_0)$-closed, then (H_C) holds.

Comments about the assumptions. By construction, Φ^* is a convex $\sigma(\mathcal{L}_0, \mathcal{U}_0)$-closed function, even if Φ is not convex. Assuming the convexity of Φ is not a restriction and this will be useful.
The assumption $(H_{\Phi 1})$ also expresses that Φ achieves its minimum at $u = 0$ and that $\Phi(0) = 0$. This is a practical normalisation requirement which will allow us to build a gauge functional associated with Φ. More, $(H_{\Phi 1})$ implies that Φ^* also shares this property: Φ^* achieves its minimum value at $\ell = 0$ and $\Phi^*(0) = 0$. Gauge functionals related to Φ^* will also appear later.
With any convex function $\tilde{\Phi}$ satisfying $(H_{\Phi 2})$, one can associate a function Φ satisfying $(H_{\Phi 1})$ in the following manner. Because of $(H_{\Phi 2})$, $\tilde{\Phi}(0)$ is finite and there exists $\ell_o \in \mathcal{L}_0$ such that $\ell_o \in \partial\tilde{\Phi}(0)$. Then, $\Phi(u) \triangleq \tilde{\Phi}(u) - \langle \ell_o, u\rangle - \tilde{\Phi}(0)$, $u \in \mathcal{U}_0$, satisfies $(H_{\Phi 1})$ and $\tilde{\Phi}^*(\ell) = \Phi^*(\ell - \ell_o) - \tilde{\Phi}(0)$, $\ell \in \mathcal{L}_0$.
The hypothesis $(H_{\Phi 3})$ is not a restriction. Indeed, let us suppose that there exists a direction $u_o \neq 0$ such that $\Phi(tu_o) = 0$ for all real t. Then any $\ell \in \mathcal{L}_0$ such that $\langle \ell, u_o\rangle \neq 0$ satisfies $\Phi^*(\ell) \geq \sup_{t \in I\!R} t\langle \ell, u_o\rangle = +\infty$ and can't be a solution to (P_0).
The hypothesis (H_{T2}) isn't a restriction either: If $y_1 - y_2 \in \ker T^T$, we have $\langle T\ell, y_1\rangle = \langle T\ell, y_2\rangle$, for all $\ell \in \mathcal{L}_0$. In other words, the spaces $\mathcal{Y}_0$ and $\mathcal{Y}_0/\ker T^T$ both specify the same constraint sets $\{\ell \in \mathcal{L}_0; T\ell = x\}$.
The effective assumptions are the following ones. $(H_{\Phi 2})$ and (H_C) are geometrical restrictions while (H_{T1}) is a regularity assumption on T. The specific form of the objective function Φ^* as a convex conjugate makes it a convex $\sigma(\mathcal{L}_0, \mathcal{U}_0)$-closed function.

3.5. **Some newcomers.** The gauge functional on $\mathcal{U}_0$ of the set $\{u \in \mathcal{U}_0; \max(\Phi(u), \Phi(-u)) \leq 1\}$ is

$$|u|_\Phi \triangleq \inf\{\alpha > 0; \max(\Phi(u/\alpha), \Phi(-u/\alpha)) \leq 1\}, u \in \mathcal{U}_0.$$

Let us define

$$\Lambda(y) \triangleq \Phi(T^T y), y \in \mathcal{Y}_0, \tag{10}$$

taking (H_{T1}) into account. The gauge functional on $\mathcal{Y}_0$ of the set $\{y \in \mathcal{Y}_0; \max(\Lambda(y), \Lambda(-y)) \leq 1\}$ is

$$|y|_\Lambda \triangleq \inf\{\alpha > 0; \max(\Lambda(y/\alpha), \Lambda(-y/\alpha)) \leq 1\}, y \in \mathcal{Y}_0.$$

It is shown at Lemma 6.2 that $|\cdot|_\Phi$ and $|\cdot|_\Lambda$ are norms. Let

$\mathcal{L} \overset{\triangle}{=} (\mathcal{U}_0, |\cdot|_\Phi)'$ be the topological dual space of $(\mathcal{U}_0, |\cdot|_\Phi)$ and let
$\mathcal{U}_1$ be the $|\cdot|_\Phi$−completion of $\mathcal{U}_0$.

Of course, we have $(\mathcal{U}_1, |\cdot|_\Phi)' = \mathcal{L} \subset \mathcal{L}_0$. Similarly, let

$\mathcal{X} \overset{\triangle}{=} \mathcal{X}_0 \cap (\mathcal{Y}_0, |\cdot|_\Lambda)'$ be the space of $|\cdot|_\Lambda$−continuous elements of $\mathcal{X}_0$ and let
$\mathcal{Y}_1$ be the $|\cdot|_\Lambda$−completion of $\mathcal{Y}_0$.

We have also $(\mathcal{Y}_1, |\cdot|_\Lambda)' = \mathcal{X} \subset \mathcal{X}_0$. We denote

$$C = C_0 \cap \mathcal{X}.$$

The $\sigma(\mathcal{U}_0, \mathcal{L}_0)$-lsc regularization of Φ is $\Phi_0^{**}(u) = \sup_{\ell\in\mathcal{L}_0}\{\langle \ell, u\rangle - \Phi^*(\ell)\}$, $u \in \mathcal{U}_0$. Of course, we have $\Phi^* = (\Phi_0^{**})^*$. We define

$$\Lambda_0(y) \overset{\triangle}{=} \Phi_0^{**}(T^T y), y \in \mathcal{Y}_0.$$

The convex $\sigma(\mathcal{U}_1, \mathcal{L})$-lsc regularization of Φ is $\Phi_1^{**}(u) = \sup_{\ell\in\mathcal{L}}\{\langle \ell, u\rangle - \Phi^*(\ell)\}, u \in \mathcal{U}_1$. It satisfies $\Phi^*(\ell) = \sup_{u\in\mathcal{U}_1}\{\langle \ell, u\rangle - \Phi_1^{**}(u)\}$, for all $\ell \in \mathcal{L}$. We define

$$\Lambda_1(y) \overset{\triangle}{=} \Phi_1^{**}(T_1^T y), y \in \mathcal{Y}_1$$

where, for any $y \in \mathcal{Y}_1$, $T_1^T y$ is the linear form on $\mathcal{L}$ defined by: $\langle T_1^T y, \ell\rangle_{\mathcal{L}^\sharp,\mathcal{L}} = \langle T\ell, y\rangle_{\mathcal{X},\mathcal{Y}_1}$, for all $\ell \in \mathcal{L}$. The above definitions of T_1^T and Λ_1 are meaningful because of the following lemma.

Lemma 3.1. *Under the hypotheses* (H_Φ) *and* (H_T)*, we have*

(a) $T\mathcal{L} \subset \mathcal{X}$
(b) $T_1^T \mathcal{Y}_1 \subset \mathcal{U}_1$.

This lemma is part of Lemma 6.2. Its proof is postponed to Section 6.7.

3.6. **The optimization problems.** The convex conjugate of Λ and Λ_0 for the duality $\langle \mathcal{Y}_0, \mathcal{X}_0\rangle$ is

$$\Lambda^*(x) \overset{\triangle}{=} \sup_{y\in\mathcal{Y}_0}\{\langle x, y\rangle - \Lambda(y)\}, x \in \mathcal{X}_0. \tag{11}$$

Let us denote $\mathcal{Y}_2 = \mathcal{X}^\sharp$ the algebraic dual space of $\mathcal{X}$ and consider

$$\bar\Lambda(y) \overset{\triangle}{=} \sup_{x\in\mathcal{X}}\{\langle x, y\rangle - \Lambda^*(x)\}, y \in \mathcal{Y}_2$$

the greatest convex $\sigma(\mathcal{Y}_2, \mathcal{X})$-lsc extension of Λ to $\mathcal{Y}_2$. This allows us to state below the *extended dual problem* (D_2). The optimization problems to be considered are

minimize $\Phi^*(\ell)$	*subject to* $T\ell \in C_0$,	$\ell \in \mathcal{L}_0$	(P_0)
minimize $\Phi^*(\ell)$	*subject to* $T\ell \in C$,	$\ell \in \mathcal{L}$	(P)
minimize $\Lambda^*(x)$	*subject to* $x \in C_0$,	$x \in \mathcal{X}_0$	$(P_0^{\mathcal{X}})$
maximize $\inf_{x\in C_0}\langle x,y\rangle - \Lambda_0(y)$,		$y \in \mathcal{Y}_0$	(D_0)
maximize $\inf_{x\in C}\langle x,y\rangle - \Lambda_1(y)$,		$y \in \mathcal{Y}_1$	(D_1)
maximize $\inf_{x\in C}\langle x,y\rangle - \bar{\Lambda}(y)$,		$y \in \mathcal{Y}_2$	(D_2)

3.7. **Statement of the abstract results.** We are now ready to give answers to the questions of Section 3.3.

Theorem 3.2 (Dual equalities and primal attainment.)**.** *Let us assume* (H_Φ) *and* (H_T).

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

$$\inf\{\Phi^*(\ell); T\ell = x, \ell \in \mathcal{L}_0\} = \Lambda^*(x) \in [0,+\infty], \forall x \in \mathcal{X}_0 \tag{12}$$

(b) *The problems* (P_0) *and* (P) *admit the same solutions and the same values.*

Let us suppose that in addition (H_C) *is fulfilled. Then,*

(c) *We have the following dual equalities*

$$\inf(P_0) = \inf(P) = \sup(D_1) = \inf_{x\in C_0} \Lambda^*(x) \in [0,+\infty].$$

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

(e) *If in addition* C_0 *is* $\sigma(\mathcal{X}_0,\mathcal{Y}_0)$*-closed, then*

$$\inf(P_0) = \sup(D_0) \in [0,+\infty]. \tag{13}$$

Let us state some remarks about these results.

- It will proved at Lemma 6.2 that dom $\Phi^* \subset \mathcal{L}$. Hence, any minimizing sequence for (P_0) stands in $\mathcal{L}$ and the topology $\sigma(\mathcal{L},\mathcal{U}_1)$ is meaningful in (d).
- Since $\mathcal{Y}_0 \subset \mathcal{Y}_1$, if C_0 is $\sigma(\mathcal{X}_0,\mathcal{Y}_0)$-closed, then C is $\sigma(\mathcal{X},\mathcal{Y}_1)$-closed.
- The little dual equality (12) is the dual equality (13) with $C_0 = \{x\}$.

We define the adjoint operator $T^* : \mathcal{Y}_2 \to \mathcal{U}_2$ for all $y \in \mathcal{Y}_2$ by:

$$\langle T^*y, \ell\rangle_{\mathcal{U}_2,\mathcal{L}} = \langle y, T\ell\rangle_{\mathcal{Y}_2,\mathcal{X}}, \forall \ell \in \mathcal{L}.$$

This definition is meaningful because $T\mathcal{L} \subset \mathcal{X}$ by Lemma 3.1.

Theorem 3.3 (Dual attainment)**.** *Let us assume* (H_Φ), (H_T) *and* (H_C). *Then, we have the following dual equality*

$$\inf(P) = \sup(D_2) \in [0,\infty] \tag{14}$$

If in addition the following constraint qualification

$$C_0 \cap \operatorname{icordom} \Lambda^* \neq \emptyset \tag{15}$$

is satisfied, then (D_2) *is attained and any solution* $\bar{y}$ *to* (D_2) *shares the following properties*

(a) $\bar{y}$ *stands in the* $\sigma(\mathcal{Y}_2,\mathcal{X})$*-closure of* dom $\Lambda \subset \mathcal{Y}_0$.

(b) $T^*\bar{y}$ *stands in the* $\sigma(\mathcal{U}_2, \mathcal{L})$*-closure of* $T^T(\mathrm{dom}\,\Lambda)$.
(c) *For any* $x_o \in \mathcal{X}$, *let us denote* $D_{x_o} = \{x \in \mathcal{X}; \Lambda^*(x_o + x) \le \Lambda^*(x_o) + 1\}$. *We also denote* $j_{D_{x_o}}$ *and* $j_{-D_{x_o}}$ *the gauge functionals on* $\mathcal{X}$ *of the convex sets* D_{x_o} *and* $-D_{x_o}$. *Then, for any* x_o *in* $C_0 \cap \mathrm{icordom}\,\Lambda^*$, $\bar{y}$ *is* $j_{D_{x_o}}$*-upper semicontinuous and* $j_{-D_{x_o}}$*-lower semicontinuous at* 0.

It will be proved at Lemma 6.2 that $\mathrm{dom}\,\Lambda^* \subset \mathcal{X}$ so that any x_o in $C_0 \cap \mathrm{icordom}\,\Lambda^*$ is in $\mathcal{X}$.

Let us denote $\mathcal{U}_2 = \mathcal{L}^\sharp$ the algebraic dual space of $\mathcal{L}$. The greatest convex $\sigma(\mathcal{U}_2, \mathcal{L})$-lsc extension of Φ is

$$\bar{\Phi}(u) \triangleq \sup_{\ell \in \mathcal{L}} \{\langle \ell, u\rangle - \Phi^*(\ell)\}, u \in \mathcal{U}_2.$$

Theorem 3.4 (Dual representation of the minimizers). *Let us assume* (H_Φ), (H_T), (H_C) *and (15). Then there exists* $(\bar{\ell}, \bar{y}) \in \mathcal{L} \times \mathcal{Y}_2$ *a solution to* (P_0) *and* (D_2).
Moreover, $(\bar{\ell}, \bar{y}) \in \mathcal{L}_0 \times \mathcal{Y}_2$ *is a solution to* (P_0) *and* (D_2) *if and only if*

(a) $\bar{x} \triangleq T\bar{\ell} \in C$
(b) $\langle \bar{y}, \bar{x}\rangle \le \langle \bar{y}, x\rangle, \forall x \in C$.
(c) $\bar{\ell} \in \partial\bar{\Phi}(T^*\bar{y})$

In this situation, $\bar{x}$ *minimizes* Λ^* *on* C_0 *and we also have* $\bar{x} \in \partial\bar{\Lambda}(\bar{y})$ *and*

$$\Phi^*(\bar{\ell}) + \bar{\Phi}(T^*\bar{y}) = \langle \bar{x}, \bar{y}\rangle = \Lambda^*(\bar{x}) + \bar{\Lambda}(\bar{y}). \tag{16}$$

In the special situation where $\mathcal{U}_0 = \mathcal{Y}_0$, $\Phi = \Lambda$ and T is the identity, the primal problem becomes $(P_0^{\mathcal{X}})$ and the above results are expressed as follows.

Corollary 3.5. *Let* $\Lambda : \mathcal{Y}_0 \mapsto [0, \infty]$ *be an extended nonnegative convex function satisfying the following requirements:* $\Lambda(0) = 0$, *and for any* $y \ne 0$, *there exist* $\alpha > 0$ *and* $t \in \mathbb{R}$ *such that* $\Lambda(\alpha y) < \infty$ *and* $\Lambda(ty) > 0$. *This is the case if* Λ *is defined by (10) and* Φ *and* T *satisfy* (H_Φ) *and* (H_T).
Let us suppose that (H_C) *is satisfied. Then,*

$$\inf_{x \in C_0} \Lambda^*(x) = \inf_{x \in C} \Lambda^*(x) = \sup(D_1) = \sup(D_2) \in [0, \infty]$$

If in addition the constraint qualification (15) holds, then there exist a solution to $(P_0^{\mathcal{X}})$ *and* (D_2). *Moreover,* $(\bar{x}, \bar{y}) \in \mathcal{X}_0 \times \mathcal{Y}_2$ *is a solution to* $(P_0^{\mathcal{X}})$ *and* (D_2) *if and only if*

(a) $\bar{x} \in C$
(b) $\langle \bar{y}, \bar{x}\rangle \le \langle \bar{y}, x\rangle, \forall x \in C$.
(c) $\bar{x} \in \partial\bar{\Lambda}(\bar{y})$

We also have

$$\langle \bar{x}, \bar{y}\rangle = \Lambda^*(\bar{x}) + \bar{\Lambda}(\bar{y}).$$

4. Some applications

In this section one gives illustrations of the general results of Section 3. Our main example is concerned with entropies. It is motivated by conditional laws of large numbers, see Section 2.1.

4.1. **Entropies.** Let $(\Omega, \mathcal{A}, R)$ be a probability space where $\mathcal{A}$ is supposed to be R-complete. The ρ^*-entropy is defined for all nonnegative measure Q on Ω, by

$$I(Q) \stackrel{\Delta}{=} \int_\Omega \rho^*(\frac{dQ}{dR})\, dR \in [0, \infty]$$

if Q is absolutely continuous with respect to the *reference probability measure* R and $I(Q) = +\infty$ otherwise.
We are concerned with the following minimization problem:

$$\textit{minimize } I(Q) \textit{ subject to } TQ \in C_0, Q \textit{ nonnegative measure on } \Omega \tag{17}$$

where T is a linear operator and C_0 is a convex set. Let us make some
Assumptions on ρ^*. *We assume that ρ^* is a closed strictly convex $[0, +\infty]$-valued function on $[0, \infty)$ which achieves its minimum at $t = 1$ with $\rho^*(1) = 0$ and which is finite on a neighbourhood of $t = 1$.*
As a consequence I uniquely achieves its minimum value at R and $I(R) = 0$. In order not to be disturbed by sign considerations, let us extend ρ^* to $\mathbb{R}$ with $\rho^*(t) = +\infty$ for any $t < 0$. It is still a closed convex function, and as such, the convex conjugate of some convex function $\rho(s)$. Let us introduce $\gamma(s) \stackrel{\Delta}{=} \rho(s) - s$, $s \in \mathbb{R}$, and the convex functional

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

where $\mathcal{U}_0$ is a space of functions on Ω to be defined later. It appears that $\gamma(s)$ is a $[0, +\infty]$-valued convex function which achieves its minimum value at 0 with $\gamma(0) = 0$ and that under some conditions to be made precise in a moment, we have

$$I(Q) = \Phi^*(Q - R) \tag{18}$$

where Φ^* is the convex conjugate of Φ for the duality $(\mathcal{L}_0, \mathcal{U}_0)$ with $\mathcal{L}_0 \stackrel{\Delta}{=} \mathcal{U}_0^\sharp$ as in the Section 3. It has been already encountered at (9). In order that the representation (18) holds, one must take $\mathcal{U}_0$ large enough to separate the measures, for instance such that the space $B(\Omega)$ of all bounded measurable functions on Ω is included in $\mathcal{U}_0$, but not too large. A good choice is

$$\mathcal{U}_0 \stackrel{\Delta}{=} \{u : \Omega \to \mathbb{R} \text{ measurable such that for all } \alpha \in \mathbb{R}, \int_\Omega \gamma(\alpha u)\, dR < \infty\}.$$

In fact, $\mathcal{U}_0$ is the Orlicz space $\mathcal{U}_0 = \mathcal{M}_{\gamma_{max}}$ associated with the Young function $\gamma_{max}(s) \stackrel{\Delta}{=} \max(\gamma(-s), \gamma(s))$ on (Ω, R). For definitions, notations and basic results on Orlicz spaces, see Section 4.2 below. It is proved in ([18], Proposition 6.2) that in this situation $\mathcal{L}$ is equal to the Orlicz space $L_{\gamma_{max}{}^*}$ and for any $\ell \in \mathcal{L}_0$,

$$\Phi^*(\ell) = \begin{cases} \int_\Omega \gamma^*(\frac{d\ell}{dR}(\omega))\, R(d\omega) & \text{if } \ell \ll R \\ +\infty & \text{otherwise} \end{cases}$$

where $\ell \ll R$ means that ℓ is a signed measure on Ω which is absolutely continuous with respect to R. Since $\gamma^*(t) = \rho^*(t + 1)$, (18) is satisfied.
One must be aware of the fact that if $\mathcal{U}_0$ had been a larger space, for instance the space of all measurable functions u on Ω such that for some $\alpha > 0$, $\int_\Omega \gamma(\alpha u)\, dR < \infty$ and $\int_\Omega \gamma(-\alpha u)\, dR < \infty$, $\mathcal{L}_0$ would have been different and Φ^* would possess a singular component: its domain would contain singular forms (in $L^s_{\gamma_{max}}$.)

To apply our previous results, one must compute the function $\bar{\Phi}$. This may be difficult in some situations. This result is stated in ([18], Theorem 6.3). Being careless with annoying details, it is essentially

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

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

$$\partial\bar{\Phi}(u) = \{\gamma'(u^{ac}).R\} \tag{19}$$

for any u in dom $\bar{\Phi}$. To make things easier, we have supposed that ρ^* is strictly convex, so that γ is differentiable. To prove (19), note that for any $h \in \mathcal{U}_0 = \mathcal{M}_{\gamma_{max}}$ and any $u \in \text{dom } \bar{\Phi} \subset L_{\gamma_{max}} \oplus L^s_{\gamma_{max}{}^*}$, we have $(u+h)^{ac} = u^{ac} + h$ and $(u+h)^s = u^s$ since $h^s = 0$. Hence, $\bar{\Phi}(u+h) - \bar{\Phi}(u) = [\int_\Omega \gamma(u^{ac}+h)\, dR + \sup\{\langle u^s, f\rangle; f \in \text{dom } \Phi^*\}] - [\int_\Omega \gamma(u^{ac})\, dR + \sup\{\langle u^s, f\rangle; f \in \text{dom } \Phi^*\}] = \int_\Omega \gamma(u^{ac}+h)\, dR - \int_\Omega \gamma(u^{ac})\, dR$.
One easily checks that our assumptions on ρ^* together with our choice of $\mathcal{U}_0$, imply that (H_Φ) is satisfied. As the transformation $\rho \to \gamma$ corresponds to the transformation $Q \to \ell = Q - R$, under the additional conditions on T and C_0 of Theorem 3.4, by this theorem the minimizer of (17) is

$$\bar{Q} = \rho'((T^*\bar{y})^{ac}) \cdot R \tag{20}$$

for some $\bar{y} \in \mathcal{Y}_2$.
Before going on with entropies, let us talk about Orlicz spaces.

4.2. **A short reminder about Orlicz spaces.** Let us recall that a Young function θ is a convex even $[0,\infty]$-valued function on $\mathbb{R}$ such that $\theta(0) = 0$. Let R be a nonnegative measure on the measure space Ω. We consider two Orlicz spaces associated with θ :

$$\mathcal{M}_\theta \triangleq \{u : \Omega \to \mathbb{R} \text{ measurable such that for all } \alpha > 0, \int_\Omega \theta(\alpha u)\, dR < \infty\}$$

and

$$\mathcal{L}_\theta \triangleq \{u : \Omega \to \mathbb{R} \text{ measurable such that for some } \alpha > 0, \int_\Omega \theta(\alpha u)\, dR < \infty\}.$$

Both are vector spaces and $\mathcal{M}_\theta \subset \mathcal{L}_\theta$ where the inclusion may be strict. We also introduce the vector spaces M_θ and L_θ which correspond to $\mathcal{M}_\theta$ and $\mathcal{L}_\theta$ when identifying R-almost equal functions.
For instance, with $\theta(s) = \theta_p(s) \triangleq |s|^p/p$, $1 < p < \infty$, we have $M_{\theta_p} = L_{\theta_p} = L_p$. But with $\theta = \tau$ given by

$$\tau(s) = e^{|s|} - |s| - 1 \tag{21}$$

provided that R is a bounded measure, we obtain

$$M_\tau = \{u; \forall\alpha > 0, \int_\Omega e^{\alpha|u|}\, dR < \infty\} \subset L_\tau = \{u; \exists\alpha > 0, \int_\Omega e^{\alpha|u|}\, dR < \infty\}$$

where the inclusion is strict if Ω contains infinitely many elements.
The gauge seminorm on $\mathcal{L}_\theta$ is $\|u\|_\theta \triangleq \inf\{\beta > 0; \int_\Omega \theta(u/\beta)\, dR \le 1\}$. It becomes a norm

on the vector spaces M_θ and L_θ which are Banach spaces. The topological dual of M_θ is L_{θ^*} for the dual bracket $\langle u, v\rangle = \int_\Omega uv\,dR$ where θ^* is the convex conjugate of θ. Note that θ^* is also a Young function.
For instance, with $\theta = \theta_p$, we have $\theta^* = \theta_q$ with $1/p + 1/q = 1$, and $L'_p = M'_\theta = L_q$, for $1 < p < \infty$. We also have $M'_\tau = L_{\tau^*}$ with

$$\tau^*(t) = (|t| + 1)\log(|t| + 1) - |t|$$

But, $L'_\theta = L_{\theta^*} \oplus L^s_\theta$ where L^s_θ is a space of singular (with respect to R) linear forms on L_θ which may not even be measures. In particular, we have $L'_\tau = L_{\tau^*} \oplus L^s_\tau$ and if $M_\theta = L_\theta$, $L^s_\theta = \{0\}$.

4.3. **Bernstein processes.** We are going to specify conditioning events for random paths and solve (17). Let $\Omega = D([0,1], S)$ be the space of cadlag paths from $[0,1]$ to a Polish space S. The state space S is endowed with its Borel σ-field $\mathcal{S}$ and Ω with $\mathcal{A}$: the R-completion of the canonical σ-field where R is some probability measure on Ω. Bernstein processes are solutions of minimization problems of the type of (17) where I is some action functional (not necessarily an entropy) and Q is subject to the marginal constraints

$$Q_a = \nu_a \text{ and } Q_b = \nu_b$$

where $0 \le a < b \le 1$, Q_t is the law of the position X_t at time t under the law Q of the process $X = (X_t)_{0\le t\le 1}$ and ν_a, ν_b are prescribed probability measures on S. We introduced a and b not to overuse the subscripts 0 and 1, but one should typically think of initial and final marginals, that is $a = 0$ and $b = 1$. It appears that a physically relevent action functional $I(Q)$ is the relative entropy $I(Q \mid R)$ with respect to R (see ?) which enters the framework of the present paper (see Section ? below).
Such an approach to the construction of Bernstein's processes is used in [12] with the relative entropy. It is based on Csiszár's results on the I-projection [9], but it unfortunately inherits a mistake of [9] (see [28], about this delicate problem).

Anyhow, it is worth (see Section 2.1) studying the minimization problem

$$\text{minimize } I(Q) \text{ subject to } Q \text{ is a probability measure, } Q_a = \nu_a \text{ and } Q_b = \nu_b \tag{22}$$

where I is a ρ^*-entropy on $\Omega = D([0,1], S)$. Taking the transformation $\rho^* \to \gamma^*$ into account, this constraint is represented by the adjoint operator $T^T(y_a, y_b) = y_a \oplus y_b$ for any $y = (y_a, y_b) \in \mathcal{Y}_0 = C_b(S)^2$ and the constraint set $C_0 = \{(\nu_a - R_a, \nu_b - R_b)\} \subset \mathcal{X}_0 = \mathcal{M}eas(S)^2$. The assumptions (H_T) and (H_C) are obviously fulfilled.
Note that $I(Q) < \infty$ implies that Q is a nonnegative measure and that the constraint $Q_a = \nu_a$ implies that $Q(\Omega) = \nu_a(S) = 1$. Therefore, under our marginal constraints any Q in dom I is a probability measure.

In order to go further, we identify the relevent spaces and topologies associated with problem (22).
We have $\mathcal{U}_0 = \mathcal{M}_{\gamma_{max}}(R)$, $\mathcal{L}_0 = \mathcal{M}_{\gamma_{max}}(R)^\sharp$, $|u|_\Phi = \|u\|_{\gamma_{max}}$, $u \in \mathcal{U}_1 = M_{\gamma_{max}}(R)$.
By a simple computation, ${\gamma_{max}}^*(\cdot) = \mathrm{cv}\min(\gamma^*(-\cdot), \gamma^*(\cdot))$ where cv stands for the convex envelope. We have $\mathcal{L} = L_{{\gamma_{max}}^*}(R)$.
Taking $\mathcal{Y}_0 = C_b(S)^2$ and $\mathcal{X}_0 = \mathcal{M}eas(S)^2$, we have $|(y_a, y_b)|_\Lambda = \|y_a \oplus y_b\|_{\gamma_{max}, R_{a,b}}$ for all $(y_a, y_b) \in \mathcal{Y}_1$ which is the space of all couples of measurable functions on S such that $y_a \oplus y_b \in L_{\gamma_{max}}(S^2, R_{a,b})$ where $R_{a,b}$ is the joint law of the couple of positions (X_a, X_b) under the process law R.

The space $\mathcal{X}$ is the set of all couples (x_a, x_b) of measures on S such that there exists $k \in L_{\gamma_{max}^*}(S^2, R_{a,b})$ with $x_a(d\alpha) = k_a(d\alpha) \triangleq [\int_S k(\alpha,\beta)\, R_{b|a}(d\beta \mid \alpha)]\, R_a(d\alpha)$ and $x_b(d\beta) = k_b(d\beta) \triangleq [\int_S k(\alpha,\beta)\, R_{a|b}(d\alpha \mid \beta)]\, R_b(d\beta)$ where $R_{a|b}(d\alpha \mid \beta)$ is the law of X_a conditionally on $X_b = \beta$ and $R_{b|a}(d\beta \mid \alpha)$ is the law of X_b conditionally on $X_a = \alpha$. We assumed that S is a Polish space to insure the existence of regular versions of conditional probability measures. The function Λ is given by

$$\Lambda(y_a, y_b) = \int_{S\times S} \gamma(y_a \oplus y_b)\, dR_{a,b}, \quad y_a, y_b \in C_b(S).$$

Taking the transformation $\rho^* \to \gamma^*$ into account together with the results of the Sections 3 and 4.1, we obtain the following

Results 1 *The problem (22) admits a solution if and only if*

$$\inf\{I(Q); Q \in \mathcal{P}(\Omega), Q_a = \nu_a, Q_b = \nu_b\} = \Psi(\nu_a, \nu_b)$$

is finite, that is if (ν_a, ν_b) is in dom Ψ *where*

$$\Psi(\nu_a, \nu_b) \triangleq \sup\{\int_S y_a\, d\nu_a + \int_S y_b\, d\nu_b - \int_{S\times S} \rho(y_a \oplus y_b)\, dR_{a,b}; y_a, y_b \in C_b(S)\}$$

This dual equality is given by Theorem 3.2. In this situation, the dual equality $\inf(P_0) = \sup(D_1)$ *yields*

$$\Psi(\nu_a, \nu_b) = \sup\{\int_S y_a\, d\nu_a \quad + \quad \int_S y_b\, d\nu_b - \int_{S\times S} \rho(y_a \oplus y_b)\, dR_{a,b};$$
$$y_a, y_b \text{ such that } y_a \oplus y_b \in M_{\gamma_{max}}(S^2, R_{a,b})\}$$

since the $\|\cdot\|_\theta$-completion of C_b is M_θ.
In the opposite direction, if the state space S is $\mathbb{R}^d$, one can choose a "small" $\mathcal{Y}_0$, for instance $\mathcal{Y}_0 = C_c^\infty(\mathbb{R}^d)^2$ which also separates $\mathcal{X}_0$, to obtain

$$\Psi(\nu_a, \nu_b) = \sup\{\int_S y_a\, d\nu_a + \int_S y_b\, d\nu_b - \int_{S\times S} \rho(y_a \oplus y_b)\, dR_{a,b}; y_a, y_b \in C_c^\infty(\mathbb{R}^d)\}$$

As I is strictly convex, by Theorem 3.2-(d), if $\Psi(\nu_a, \nu_b) < \infty$, any minimizing sequence of (22) $\sigma(L_{\gamma_{max}^*}(R), M_{\gamma_{max}}(R))$-converges to the unique minimizer $\bar{Q}$.
We also have

$$I(\bar{Q}) = \Psi(\nu_a, \nu_b).$$

If (ν_a, ν_b) is in icordom Ψ, one can apply Theorems 3.3 and 3.4 to obtain the representation (20) of the minimizer. Our aim now is to give some details about the dual parameter $T^*\bar{y}$. By Theorem 3.3, $T^*\bar{y} \in$ dom $\bar{\Phi}$ stands in the $\sigma(\mathcal{U}_2, \mathcal{L})$-closure of $T^T(\text{dom}\,\Lambda)$. By (20), one only needs to consider its absolutely continuous part $(T^*\bar{y})^{ac}$. Because of the special form of $\bar{\Phi}$, we know that the absolutely continuous parts of the elements of dom $\bar{\Phi}$ are in $L_{\gamma_{max}}(R)$ and a fortiori $(T^*\bar{y})^{ac} \in L_1(R)$. We have just seen that $\mathcal{L} = L_{\gamma_{max}^*}(R)$. Therefore, $L_\infty(R) \subset \mathcal{L}$. As any singular form in $L^s_{\gamma_{max}^*}$ vanishes at any bounded function, it follows that $(T^*\bar{y})^{ac}$ is in the $\sigma(L_1, L_\infty)$-closure of the convex set $T^T(\text{dom}\,\Lambda)$ which is also the $\|\cdot\|_1$-closure of $T^T(\text{dom}\,\Lambda) = \{y_a \oplus y_b; \int_{S^2} \gamma(y_a \oplus y_b)\, dR_{a,b} < \infty\}$. Hence, one can extract a sequence (y_a^n, y_b^n) in dom Λ such that $y_a^n(X_a) \oplus y_b^n(X_b)$ converges R-almost everywhere to $(T^*\bar{y})^{ac}$. This implies that $y_a^n \oplus y_b^n$ converges $R_{a,b}$-almost everywhere to some function $h \in L_{\gamma_{max}}(S^2, R_{a,b})$. By ([4], Corollary 3.4) this limit h keeps the same additive form: $h = \bar{y}_a \oplus \bar{y}_b$ $R_{a,b}$-almost everywhere. Hence, there exist functions $\bar{y}_a$ and $\bar{y}_b$ on S such that

$$(T^*\bar{y})^{ac} = \bar{y}_a(X_a) + \bar{y}_b(X_b), R\text{−almost everywhere}$$

Note that it is not clear that, although h is jointly measurable, both $\bar{y}_a$ and $\bar{y}_b$ are measurable. One concludes that this holds in some sense, thanks to Lemma 4.1 below. We have obtained the following

Results 2 *If (ν_a, ν_b) stands in* icordom Ψ, *there exist two numerical functions $\bar{y}_a$ and $\bar{y}_b$ on S such that $\bar{y}_a \oplus \bar{y}_b$ is measurable with respect to the $R_{a,b}$-completion of $\mathcal{S}^{\otimes 2}$ and the unique minimizer $\bar{Q}$ of Problem (22) is*

$$\bar{Q}(d\omega) = \rho'(\bar{y}_a(\omega_a) + \bar{y}_b(\omega_b)) \cdot R(d\omega) \tag{23}$$

For R_b-almost every $\beta \in S$, $\bar{y}_a$ is measurable with respect to the $R_{a|b}(\cdot|\beta)$-completion of $\mathcal{S}$ and for R_a-almost every $\alpha \in S$, $\bar{y}_b$ is measurable with respect to the $R_{b|a}(\cdot|\alpha)$-completion of $\mathcal{S}$.
We also have:

$$\int_S \bar{y}_a \, d\nu_a + \int_S \bar{y}_b \, d\nu_b = \int_{S^2} \rho(\bar{y}_a(\alpha) + \bar{y}_b(\beta)) \, R_{a,b}(d\alpha d\beta) + \Psi(\nu_a, \nu_b)$$

Lemma 4.1 (Negligible troubles)**.** *Let us consider A and B two Polish spaces endowed with their Borel σ-fields $\mathcal{A}$ and $\mathcal{B}$, the product space $A \times B$, its product σ-field $\mathcal{A} \otimes \mathcal{B}$, $X = (X_A, X_B)$ the canonical couple on $A \times B$ and P a probability measure on $A \times B$ interpreted as the law of X. We endow $A \times B$ with the P-completion of $\mathcal{A} \otimes \mathcal{B}$ and extend P to this σ-field. Let us denote P_A, P_B, $P_{A|X_B=b}$ and $P_{B|X_A=a}$ the laws of X_A, X_B and the regular conditional versions of the laws of X_A conditioned on X_B and X_B conditioned on X_A.*
If $h : A \times B \to I\!R$ is a jointly measurable function such that $h = f(X_A) + g(X_B)$, P-almost everywhere, then for P_B-almost every b, $f : A \to I\!R$ is measurable with respect to the $P_{A|X_B=b}$-completion of $\mathcal{A}$ and for P_A-almost every a, $g : B \to I\!R$ is measurable with respect to the $P_{B|X_A=a}$-completion of $\mathcal{B}$.
In particular, if X_A and X_B are P-independent, then f is measurable with respect to the P_A-completion of $\mathcal{A}$ and g is measurable with respect to the P_B-completion of $\mathcal{B}$.

Proof. Let h be a measurable function on $A \times B$ and k a not necessarily measurable function such that $k(X) = h(X)$, P-almost surely. There exists a measurable set N such that $P(N) = 0$ and $k(X) = h(X)$ on $(A \times B) \setminus N$. Under the Polish assumption, one can take advantage of the existence of regular versions of the conditional expectations. As $P(N) = \int_B P(N \mid X_B = b) \, P_B(db) = 0$, we obtain that P_B-almost everywhere, $P(N \mid X_B = b) = 0$. Let $N_{(X_B=b)} = \{a \in A; (a,b) \in N\}$ be the $(X_B = b)$-section of N. Since we work with the P-completion of $\mathcal{A} \otimes \mathcal{B}$, the section $N_{(X_B=b)}$ is measurable with respect to the $P_{A|X_B=b}$-completion of $\mathcal{A}$. Therefore, we have $P_A(N_{(X_B=b)} \mid X_B = b) = P(N \mid X_B = b) = 0$ and $k(a,b) = h(a,b)$ for P_B-almost every b and all $a \notin N_{(X_B=b)}$.
As h is jointly measurable, for every $b_o \in B$, the section $a \mapsto h(a, b_o)$ is measurable with

respect to the $P_{A|X_B=b_o}$-completion of $\mathcal{A}$. Taking $k(a,b) = f(a)+g(b)$ in the above result, we obtain that $a \mapsto f(a)+g(b_o)$ is measurable with respect to the $P_{A|X_B=b_o}$-completion of $\mathcal{A}$, for P_B-almost every b_o. This proves the result for f. A similar proof works for g. The last statement of the lemma follows from $P_{A|X_B=b} = P_A$ for P_B-almost every b and $P_{B|X_A=a} = P_B$ for P_A-almost every a, in the case where $P = P_A \otimes P_B$. □

4.4. **Nelson processes.** Nelson processes are solutions of minimization problems of the type of (17) where I is some action functional (not necessarily an entropy) and Q is subject to the marginal constraints

$$Q_t = \nu_t, \forall 0 \leq t \leq 1$$

where $\nu = (\nu_t)_{0\leq t\leq 1}$ is a prescribed flow of probability measures on S. The minimization problem to be considered in this paper is similar to Bernstein's one:

$$\textit{minimize } I(Q) \textit{ subject to } Q \textit{ is a probability measure, } Q_t = \nu_t \textit{ for all } \ 0 \leq t \leq 1 \qquad (24)$$

where I is a ρ^*-entropy on $\Omega = D([0,1],S)$ as in Section 4.3.

We assume that $t \mapsto \nu_t$ is continuous. Without loss of information, we identify the flow ν with the probability measure $\nu_t(d\alpha)dt$ on $[0,1]\times S$. The centering procedure is now $x_t(d\alpha)dt = [\nu_t - R_t](d\alpha)dt$, and the dual bracket $\langle x,y\rangle$ is given by $\langle x,y\rangle = \int_{[0,1]\times S} y(t,\alpha)x_t(d\alpha)dt$ for all $y \in \mathcal{Y}_0 \triangleq C_b([0,1]\times S)$. The constraint is expressed for all $y \in C_b([0,1]\times S)$ by

$$T^T y = \int_{[0,1]} y(t,X_t)\,dt$$

and the function Λ is

$$\Lambda(y) = \int_\Omega \left(\int_{[0,1]} \gamma[y(t,\omega_t)]\,dt\right) R(d\omega).$$

Following the approach of the previous section, we obtain the following

Results 3 *The problem (24) admits a solution if and only if*

$$\inf\{I(Q); Q\in\mathcal{P}(\Omega), Q_t=\nu_t, \forall 0\leq t\leq 1\} = \Psi(\nu)$$

is finite, where

$$\Psi(\nu) \triangleq \sup\{\int_{[0,1]\times S} y(t,\alpha)\,\nu_t(d\alpha)dt - \int_\Omega \rho(y(t,\omega_t))\,R(d\omega); y \in C_b([0,1]\times S)\}$$

One also gets

$$\Psi(\nu) = \sup\{\int_{[0,1]\times S} y(t,\alpha)\,\nu_t(d\alpha)dt - \int_\Omega \rho(y(t,\omega_t))\,R(d\omega); y \in M_{\gamma_{max}}([0,1]\times S, R_t(d\alpha)dt)\}. \qquad (25)$$

Any minimizing sequence of (24) $\sigma(L_{\gamma_{max}{}^*}(R), M_{\gamma_{max}}(R))$-converges to the unique minimizer $\bar{Q}$. We also have $I(\bar{Q}) = \Psi(\nu)$.

If ν stands in icordom Ψ, *there exists a measurable function $\bar{y}$ on $[0,1]\times S$ in $L_{\gamma_{max}}([0,1]\times S, R_t(d\alpha)dt)$ such that the unique minimizer $\bar{Q}$ of Problem (24) is given by*

$$\bar{Q}(d\omega) = \rho'\left(\int_{[0,1]} \bar{y}(t,\omega_t)\,dt\right)\cdot R(d\omega)$$

We also have:

$$\int_{[0,1]\times S} \bar{y}(t,\alpha)\,\nu_t(d\alpha)dt = \int_{[0,1]\times S} \rho(\bar{y}(t,\alpha))\,R_t(d\alpha)dt + \Psi(\nu)$$

4.5. **Relative entropy and conditional independence.** As already remarked the relative entropy of a probability Q on Ω with respect to R

$$I(Q\mid R) \triangleq \begin{cases} \int_\Omega \log\left(\frac{dQ}{dR}\right)\,dQ & \text{if } Q\in\mathcal{P}(\Omega) \text{ and } Q\ll R \\ +\infty & \text{otherwise}\end{cases}$$

is a ρ^*-entropy with $\rho^*(t)=t\log t - t + 1$, $t\geq 0$ ($\rho^*(0)=1$) and the constraint $Q(\Omega)=1$. This corresponds to $\rho(s)=e^s-1$, $\gamma_{max}(s)=e^{|s|}-|s|-1$ and the representation (20) becomes

$$\bar{Q} = \exp\left((T^*\bar{y})^{ac}\right)\cdot R$$

Roughly speaking, if R has a product form and $T^*\bar{y}$ has a sum form, $\bar{Q}$ preserves the product form. We are going to formalize this remark. This will lead us to the following typical result: if R is the law of a *Markov* process and $\bar{Q}$ is the solution of the minimization problem

$$\textit{minimize } I(Q\mid R) \textit{ subject to } Q\in\mathcal{P}(\Omega), TQ\in C_o \tag{26}$$

where $TQ\in C_o$ is a convex marginal constraint (Bernstein or Nelson processes, for instance), then $\bar{Q}$ is also the law of a Markov process. Note that the mass constraint $Q(\Omega)=1$ should enter the definition of T when considering $I(Q\mid R)$ as a ρ^*-entropy.
The spaces $M_{\gamma_{max}}$ and $L_{\gamma_{max}}$ are

$$\begin{aligned} M_\tau(R) &\triangleq \{u:\Omega\to I\!R;\ \text{measurable}, \forall\alpha>0, \int_\Omega e^{\alpha|u|}\,dR<\infty\} \\ L_\tau(R) &\triangleq \{u:\Omega\to I\!R;\ \text{measurable}, \exists\alpha>0, \int_\Omega e^{\alpha|u|}\,dR<\infty\}\end{aligned}$$

since $\gamma_{max}(s)=\tau(s)\triangleq e^{|s|}-|s|-1$. Recall that $\tau^*(t)=(|t|+1)\log(|t|+1)-|t|$. Any minimizing sequence of (26) is $\sigma(L_{\tau^*},M_\tau)$-converging to $\bar{Q}$.

Let's have a look at T. Let $\theta:\Omega\to\mathcal{X}_0$ be a measurable function, in the sense that for all $y\in\mathcal{Y}_0$, $\omega\in\Omega\mapsto\langle y,\theta(\omega)\rangle\in I\!R$ is a measurable function. The operator T is defined by $T^Ty\triangleq\langle y,\theta(\cdot)\rangle$, for all $y\in\mathcal{Y}_0$. This means that

$$\langle y, TQ\rangle = \int_\Omega \langle y,\theta(\omega)\rangle\,Q(d\omega).$$

The marginal constraint of Bernstein processes corresponds to $\theta(\omega)=(\delta_{\omega_a},\delta_{\omega_b})\in\mathcal{X}_0=\mathcal{M}eas(S)^2$ and the marginal constraint of Nelson processes corresponds to $\theta(\omega)=\delta_{\omega_t}(d\alpha)dt\in\mathcal{X}_0=\mathcal{P}([0,1]\times S)$.
In terms of θ, the assumption $(H_T)$is

(1) $\int_\Omega \exp(|\langle y, \theta(\omega)\rangle|)\, R(d\omega) < \infty$, for all $y \in \mathcal{Y}_0$.
(2) For $y \in \mathcal{Y}_0$, if $\langle y, \theta(\omega)\rangle = 0, \forall \omega \in \Omega$, then $y = 0$.

These requirements are assumed to hold together with the assumption (H_C) on C_o.
The little dual equality is

$$\inf\{I(Q \mid R); TQ = z\} = \Psi(z) \triangleq \sup_{y \in \mathcal{Y}_0}\{\langle y, z\rangle - \int_\Omega e^{\langle y, \theta(\omega)\rangle}\, R(d\omega)\} + 1$$

To be able to state our result about conditional independence and relative entropy, one must introduce

Additional assumptions on θ and R. *There exist three measure spaces $(\Omega_k, \mathcal{A}_k)$, $k = a, b, c$, a measurable function $\alpha = (\alpha_a, \alpha_b, \alpha_c) : (\Omega, \mathcal{A}) \to (\Omega_a \times \Omega_b \times \Omega_c, \mathcal{A}_a \otimes \mathcal{A}_b \otimes \mathcal{A}_c)$ and three measurable functions $\theta_k : (\Omega_k, \mathcal{A}_k) \to \mathcal{X}_0$, $k = a, b, c$, in the sense that for all $y \in \mathcal{Y}_0$, $\langle y, \theta_k\rangle$ is measurable, such that*

$$\theta = \theta_a \circ \alpha_a + \theta_b \circ \alpha_b + \theta_c \circ \alpha_c \tag{27}$$

We also assume that α_a and α_b are R-independent conditionally on α_c. This means that for every $A_k \in \mathcal{A}_k$, $k = a, b, c$

$$R((\alpha_a, \alpha_b) \in A_a \times A_b \mid \alpha_c \in A_c) = R(\alpha_a \in A_a \mid \alpha_c \in A_c) R(\alpha_b \in A_b \mid \alpha_c \in A_c). \tag{28}$$

An example: Markov processes. Let $\Omega = S^{[0,1]}$ be endowed with the natural σ-field generated by the the canonical process $X = (X_t)_{0 \le t \le 1}$. For any $0 \le t \le 1$ let $\mathcal{P}_t = \sigma(X_s; 0 \le s \le t)$, $\mathcal{F}_t = \sigma(X_s; t \le s \le 1)$ and $\mathcal{N}_t = \sigma(X_t)$ be the σ-field of the past, future and present (now) at time t. The law R of X is Markov if and only if for any $0 < t < 1$, $\mathcal{P}_t$ and $\mathcal{F}_t$ are $R(\cdot \mid \mathcal{N}_t)$-independent. This fits (28) choosing

- $\Omega_a = S^{[0,t]}$, $\mathcal{A}_a = \mathcal{P}_t$, $\alpha_a(\omega) = (\omega_s)_{0 \le s \le t}$
- $\Omega_b = S^{[t,1]}$, $\mathcal{A}_b = \mathcal{F}_t$, $\alpha_b(\omega) = (\omega_s)_{t \le s \le 1}$
- $\Omega_c = S^{\{t\}}$, $\mathcal{A}_c = \mathcal{N}_t$, $\alpha_c(\omega) = \omega_t$.

Bernstein's marginal constraints. We consider the constraints $Q_0 = \nu_0$ and $Q_1 = \nu_1$: that is $\theta = (\delta_{X_0}, \delta_{X_1}) \in \mathcal{M}eas(S)^2$. Choosing $\theta_a \circ \alpha_a = (\delta_{X_0}, 0)$, $\theta_b \circ \alpha_b = (0, \delta_{X_1})$ and $\theta_c \circ \alpha_c = (0, 0)$, one obtains the representation formula (27).
Nelson's marginal constraints. We consider the constraints $Q_t = \nu_t$, for all $0 \le t \le 1$: that is $\theta(\omega)(dsd\alpha) = \delta_{\omega_s}(d\alpha)ds \in \mathcal{M}eas([0,1] \times S)$. Choosing $\theta_a \circ \alpha_a(\omega)(dsd\alpha) = \mathbf{1}_{[0,t]}(s)\delta_{\omega_s}(d\alpha)ds$, $\theta_b \circ \alpha_b(\omega)(dsd\alpha) = \mathbf{1}_{[t,1]}(s)\delta_{\omega_s}(d\alpha)ds$, $\theta_c \circ \alpha_c = 0$, one obtains the representation formula (27).

Theorem 4.2. *We suppose that Ω is a Polish space endowed with the R-completion of its Borel σ-field. We assume that* icordom *$\Psi \cap C_o$ is nonempty. Under the above assumptions on θ and R, the minimizer $\bar{Q}$ of (17) shares the same conditional independence property (28) as R :*

$$\bar{Q}((\alpha_a, \alpha_b) \in A_a \times A_b \mid \alpha_c \in A_c) = \bar{Q}(\alpha_a \in A_a \mid \alpha_c \in A_c)\bar{Q}(\alpha_b \in A_b \mid \alpha_c \in A_c)$$

for every $A_k \in \mathcal{A}_k$, $k = a, b, c$.
We also have

$$\bar{Q}(d\omega) = \exp\langle \bar{y}, \theta(\omega)\rangle \cdot R(d\omega)$$

for some linear form $\bar{y}$ on $\mathcal{X}_0$ such that $\langle \bar{y}, \theta\rangle \in L_\tau(R)$.

Proof. We shall make use of the following basic result. If $P = ZR$, Y is a bounded random variable and $\mathcal{H}$ is a σ-field,

$$E_P(Y \mid \mathcal{H}) E_R(Z \mid \mathcal{H}) = E_R(ZY \mid \mathcal{H}).$$

The last identity of the theorem is a consequence of Theorems 3.3 and 3.4. As one proves it similarly as (23), we omit the details.
Therefore, $Z \stackrel{\Delta}{=} \frac{d\bar{Q}}{dR} = Z_a Z_b Z_c$ with $Z_k = \exp\langle \bar{y}, \theta_k \rangle$, $k = a, b, c$. Thanks to the last statement of Lemma 4.1 with the conditional independence property (28), we obtain that Z_a is measurable with respect to the $R_a(\cdot \mid \mathcal{A}_c)$-completion of $\mathcal{A}_a$ where $R_a \stackrel{\Delta}{=} R \circ \alpha_a^{-1}$ and Z_b is measurable with respect to the $R_b(\cdot \mid \mathcal{A}_c)$-completion of $\mathcal{A}_b$ where $R_b \stackrel{\Delta}{=} R \circ \alpha_b^{-1}$. It follows that for any Y_a and Y_b bounded and measurable with respect to $\mathcal{A}_a$ and $\mathcal{A}_b$ respectively,

$$E_{\bar{Q}}(Y_a \mid \mathcal{A}_c) = \frac{E_R(Y_a Z_a \mid \mathcal{A}_c)}{E_R(Z_a \mid \mathcal{A}_c)} \quad \text{and} \quad E_{\bar{Q}}(Y_b \mid \mathcal{A}_c) = \frac{E_R(Y_b Z_b \mid \mathcal{A}_c)}{E_R(Z_b \mid \mathcal{A}_c)}.$$

are well defined. Then $E_{\bar{Q}}(Y_a Y_b \mid \mathcal{A}_c) = \frac{E_R(Y_a Z_a Y_b Z_b Z_c \mid \mathcal{A}_c)}{E_R(Z_a Z_b Z_c \mid \mathcal{A}_c)} = \frac{E_R(Y_a Z_a \mid \mathcal{A}_c)}{E_R(Z_a \mid \mathcal{A}_c)} \frac{E_R(Y_b Z_b \mid \mathcal{A}_c)}{E_R(Z_b \mid \mathcal{A}_c)}$, by conditional independence. Hence, $E_{\bar{Q}}(Y_a Y_b \mid \mathcal{A}_c) = E_{\bar{Q}}(Y_a \mid \mathcal{A}_c) E_{\bar{Q}}(Y_b \mid \mathcal{A}_c)$ which is the desired result. □

5. Convex minimization

For the convenience of the reader, we are going to sketch the main lines of the modern approach to convex minimization problems, by means of conjugate duality, as developed in Rockafellar's monograph [27].

Let A be a vector space and $f : A \to [-\infty, +\infty]$ an extended real convex function. We consider the following convex minimization problem

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

Let Q be another vector space. The *pertubation* of the *objective function* f is a function $F : A \times Q \to [-\infty, +\infty]$ such that for $q = 0 \in Q$, $F(\cdot, 0) = f(\cdot)$. The problem $(\mathcal{P})$ is imbedded in a parametrized family of minimization problems

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

The *value function* of $(\mathcal{P}_q)_{q \in Q}$ is

$$\varphi(q) \stackrel{\Delta}{=} \inf(\mathcal{P}_q) = \inf_{a \in A} F(a, q) \in [-\infty, +\infty], q \in Q.$$

Let us assume that the perturbation is chosen such that

$$F \text{ is jointly convex on } A \times Q. \tag{29}$$

Then, $(\mathcal{P}_q)_{q \in Q}$ is a family of convex minimization problems. More, the value function φ is convex. This follows from the fact that the epigraph of φ is "essentially" a linear (marginal) projection of the convex epigraph of F so that it is also convex.

Let B be a vector space in separating duality with Q. The *Lagrangian* associated with the perturbation F and the duality $\langle B, Q \rangle$ is

$$K(a, b) \stackrel{\Delta}{=} \inf_{q \in Q} \{\langle b, q \rangle + F(a, q)\}, a \in A, b \in B. \tag{30}$$

In other words, for any $a \in A$, $b \mapsto K(a,b)$ is the concave conjugate of the function $q \mapsto -F(a,q)$ and as such it is a concave function. Under (29), for any $b \in B$, $a \mapsto K(a,b)$ is a convex function (same argument as for the convexity of φ.) Therefore, K is a *convex-concave* function. We shall see that its *saddle-points* will play a central role.

Let us assume that $\langle B, Q\rangle$ is a *dual pairing:* this means that B and Q are locally convex topological vector spaces such that their topological dual spaces B' and Q' satisfy $B' = Q$ and $Q' = B$ up to some isomorphisms. A typical instance of dual pairing is given by the weak topologies $\sigma(Q,B)$ and $\sigma(B,Q)$. More generally: Q is a locally convex topological vector space and $B = Q'$ is equipped with $\sigma(B,Q)$. An interesting aspect of dual pairing is that, thanks to Hahn-Banach's theorem, the closed convex sets are the weakly closed convex sets. Applied to the epigraphs, this implies that closed convex functions are weakly closed convex functions.

Assuming in addition that F is chosen such that

$$q \mapsto F(a,q) \text{ is a closed convex function for any } a \in A, \tag{31}$$

one can reverse the conjugate duality relation (30) to obtain

$$F(a,q) = \sup_{b \in B}\{K(a,b) - \langle b, q\rangle\}, \forall a \in A, q \in Q \tag{32}$$

Let us think of K as a pivot: $-K$ is concave in a and convex in b. This suggests to introduce another vector space P in separating duality with A such that $\langle P, A\rangle$ is a dual pairing and to introduce also the function

$$G(b,p) \triangleq \inf_{a \in A}\{K(a,b) - \langle a, p\rangle\}, b \in B, p \in P. \tag{33}$$

This formula is analogous to (32). Since

$$G(b,p) = \inf_{a,q}\{\langle b, q\rangle - \langle a, p\rangle + F(a,q)\}, b \in B, p \in P,$$

one sees that G is jointly closed concave, as a concave conjugate. Going on symetrically, one interprets G as the concave perturbation of the objective concave function

$$g(b) \triangleq G(b,0), b \in B$$

associated with the concave maximization problem

$$\textit{maximize } g(b), b \in B \tag{$\mathcal{D}$}$$

which is called the *dual problem* of the *primal problem* $(\mathcal{P})$. It is imbedded in the family of concave maximization problems $(\mathcal{D}_p)_{p \in P}$

$$\textit{maximize } G(b,p), b \in B \tag{$\mathcal{D}_p$}$$

whose value function is

$$\gamma(p) \triangleq \sup_{b \in B} G(b,p), p \in P.$$

Since G is jointly concave, γ is also concave. We have the following diagram

$$\begin{array}{ccccccc} & & \gamma(p) & & f(a) & & \\ & \Big\langle & P & , & A & \Big\rangle & \\ G(b,p) & & & K(a,b) & & & F(a,q) \\ & \Big\langle & B & , & Q & \Big\rangle & \\ & & g(b) & & \varphi(q) & & \end{array}$$

Because of (32) and (33) with $q=0$ and $p=0$ we obtain

$$f(a) = \sup_{b\in B} K(a,b), a \in A \tag{34}$$

$$g(b) = \inf_{a\in A} K(a,b), b \in B \tag{35}$$

and the values of the optimization problems satisfy

$$\sup(\mathcal{D}) = \gamma(0) = \sup_b g(b) = \sup_b \inf_a K(a,b) \le \inf_a \sup_b K(a,b) = \inf_a f(a) = \varphi(0) = \inf(\mathcal{P}).$$

It appears that the *dual equality:* $\inf(\mathcal{P}) = \sup(\mathcal{D})$ holds if and only if K has a *saddle-value.*

Maybe the main result of this theory is the following one.

Lemma 5.1. *We assume that $\langle P, A\rangle$ and $\langle B, Q\rangle$ are dual pairings.*

(a) *Without any additional assumptions, we have*

$$-g(-b) = \varphi^*(b), \forall b \in B.$$

(b) *Under the additional assumption (31), (29) is equivalent to: $a \mapsto K(a,b)$ is closed convex for all $b \in B$.*

(c) *Under the additional assumptions (29) and (31), we have*

$$f(a) = (-\gamma)^*(a), \forall a \in A.$$

Proof. The proof of (b) is left to the reader (see [27]). Statement (a) is a direct consequence of the definitions. Indeed, for all $b \in B$, $g(b) \overset{\Delta}{=} \inf_a K(a,b) \overset{\Delta}{=} \inf_{a,q}\{\langle b,q\rangle + F(a,q)\} \overset{\Delta}{=} \inf_q\{\langle b,q\rangle + \varphi(q)\}$, which is the desired result.
Let us prove (c). Statement (b) allows us to reverse the conjugate duality relation (33) to obtain

$$K(a,b) = \sup_{p\in P}\{\langle a,p\rangle + G(b,p)\}, \forall a \in A, b \in B.$$

Therefore, for all $a \in A$, $f(a) = \sup_b K(a,b) = \sup_{b,p}\{\langle a,p\rangle + G(b,p)\} \overset{\Delta}{=} \sup_p\{\langle a,p\rangle + \gamma(p)\}$, which is the desired result. The first equality is (34) and the last one is the definition of γ. $\square$

Note that by (c), under our assumptions (29) and (31), f must be a closed convex function.

Theorem 5.2. *We assume that $\langle P, A\rangle$ and $\langle B, Q\rangle$ are dual pairings.*

(a) *We have* $\sup(\mathcal{D}) = \mathrm{cl\ co}\ \varphi(0)$, *where* $\mathrm{cl\ co}\ \varphi$ *is the closed convex regularization of* φ. *The dual equality* $\inf(\mathcal{P}) = \sup(\mathcal{D})$ *holds if and only if* $\varphi(0) = \mathrm{cl\ co}\ \varphi(0)$. *In particular, this is the case if (29) holds and φ is lower semicontinuous at $q = 0$.*

(b) *If the dual equality holds, then*

$$\operatorname{argmax} g = -\partial\varphi(0) \triangleq \{b \in B; \varphi(q) - \varphi(0) \geq -\langle b, q\rangle, \forall q \in Q\}$$

where argmax g *is the set of maximizers of* $(\mathcal{D})$ *and* $\partial\varphi(0)$ *is the subdifferential of* φ *at 0, for the duality* $\langle B, Q\rangle$.

Let us assume in addition that (29) and (31) are satisfied.

(c) *We have* $\inf(\mathcal{P}) = -\mathrm{cl}\,(-\gamma)(0)$ *where* $\mathrm{cl}\,(-\gamma)$ *is the closed regularization of the convex function* $-\gamma$. *The dual equality* $\inf(\mathcal{P}) = \sup(\mathcal{D})$ *holds if and only if* $-\gamma(0) = \mathrm{cl}\,(-\gamma)(0)$. *In particular, this is the case if* γ *is upper semicontinuous at* $p = 0$.

(d) *If the dual equality holds, then*

$$\operatorname{argmin} f = \partial(-\gamma)(0) \triangleq \{a \in A; -\gamma(p) + \gamma(0) \geq \langle a, p\rangle, \forall p \in P\}$$

where argmin f *is the set of minimizers of* $(\mathcal{P})$ *and* $\partial(-\gamma)(0)$ *is the subdifferential of* $-\gamma$ *at 0, for the duality* $\langle P, A\rangle$.

Proof. Let us prove (a). Reversing the conjugate identity of Lemma 5.1-(a), one obtains for all $q \in Q$, $\varphi^{**}(q) = \sup_b\{\langle b, q\rangle + g(-b)\}$. In particular, with $q = 0$, one gets $\varphi^{**}(0) = \sup_b g(b) = \sup(\mathcal{D})$. The dual equality is $\varphi(0) = \varphi^{**}(0)$. It is the desired result since the biconjugate φ^{**} is the closed convex regularization of φ.
Let us prove (b). Because of the conjugate identity of Lemma 5.1-(a), since the dual equality is $\varphi(0) = \varphi^{**}(0)$, we have for any $b \in B$, $\partial(-g)(-b) = \partial\varphi^*(b) \ni 0$ if and only if $b \in \partial\varphi(0)$. This implies that $\bar{b} \in \operatorname{argmax} g$ if and only if $-\bar{b} \in \partial\varphi(0)$, which is the desired result.
Let us prove (c). Reversing the conjugate identity of Lemma 5.1-(c), one obtains for all $p \in P$, $(-\gamma)^{**}(p) = f^*(p) = \sup_a\{\langle a, p\rangle - f(a)\}$. In particular, with $p = 0$, we get $-(-\gamma)^{**}(0) = \inf_a f(a) = \inf(\mathcal{P})$ and the dual equality is $-\gamma(0) = (-\gamma)^{**}(0)$.
Let us prove (d). As in (b), since $-\gamma(0) = (-\gamma)^{**}(0)$, $\partial f(\bar{a}) = \partial(-\gamma)^*(\bar{a}) \ni 0$ is equivalent to $\bar{a} \in \partial(-\gamma)(0)$. □

Let us investigate now what is usually called the *Kuhn-Tucker conditions.* It is related to the notion of saddle-points of the Lagrangian K. Suppose that the optimization problems $(\mathcal{P})$ and $(\mathcal{D})$ are both attained. This means that there exist $\bar{a} \in A$ and $\bar{b} \in B$ such that for all $a \in A$, $f(\bar{a}) \leq f(a)$ and all $b \in B$, $g(\bar{b}) \geq g(b)$. Because of (34) and (35), we obtain for all a and b that $g(b) \leq K(a, b) \leq f(a)$. Suppose that the dual equality $f(\bar{a}) = g(\bar{b})$ also holds. Then, we have $g(\bar{b}) = K(\bar{a}, \bar{b}) = f(\bar{a})$ and

$$K(\bar{a}, b) \leq K(\bar{a}, \bar{b}) \leq K(a, \bar{b}), \forall a \in A, b \in B.$$

This means that $(\bar{a}, \bar{b})$ is a saddle-point of K. Hence, we have proved the following important result.

Theorem 5.3 (Kuhn-Tucker conditions)**.** *We assume that* $\langle P, A\rangle$ *and* $\langle B, Q\rangle$ *are dual pairings and that (29) and (31) are satisfied. Suppose that* $(\mathcal{P})$ *is attained at* $\bar{a} \in A$, $(\mathcal{D})$ *is attained at* $\bar{b} \in B$ *and the dual equality holds. Then we have*

$$\partial_a K(\bar{a}, \bar{b}) \ni 0 \tag{36}$$

$$\partial_b(-K)(\bar{a}, \bar{b}) \ni 0 \tag{37}$$

where $\partial_a K(\bar a,\bar b) \triangleq \{p\in P; K(a,\bar b)-K(\bar a,\bar b)\geq \langle p, a-\bar a\rangle, \forall a\in A\}$ *and* $\partial_b(-K)(\bar a,\bar b) \triangleq \{q\in Q; -K(\bar a,b)+K(\bar a,\bar b)\geq \langle q, b-\bar b\rangle, \forall b\in B\}$ *are the subdifferentials of the convex functions* $a\mapsto K(a,\bar b)$ *and* $b\mapsto -K(\bar a,b)$ *at* $\bar a$ *and* $\bar b$ *respectively.*

6. The proofs of the results of Section 3

We are going to apply the general results of the Lagrangian approach of Section 5 to the minimization problem (P_0).

6.1. **A first dual equality.** Let us begin applying the Lagrangian approach with $\langle P,A\rangle = \langle \mathcal{U}_0,\mathcal{L}_0\rangle$ and $\langle B,Q\rangle = \langle \mathcal{Y}_0,\mathcal{X}_0\rangle$ and the topologies are the weak topologies $\sigma(\mathcal{L}_0,\mathcal{U}_0)$, $\sigma(\mathcal{U}_0,\mathcal{L}_0)$, $\sigma(\mathcal{X}_0,\mathcal{Y}_0)$ and $\sigma(\mathcal{Y}_0,\mathcal{X}_0)$. The function to be minimized is $f_0(\ell)=\Phi^*(\ell)+\delta_{C_0}(T\ell)$, $\ell\in\mathcal{L}_0$ where $\delta_C(x)=\begin{cases}0 & \text{if } x\in C\\ +\infty & \text{if } x\notin C\end{cases}$ denotes the convex indicator of C. The perturbation F_0 of f_0 is Fenchel's one:

$$F_0(\ell,x)=\Phi^*(\ell)+\delta_{C_0}(T\ell+x), \ell\in\mathcal{L}_0, x\in\mathcal{X}_0$$

the value function of which is

$$\varphi_0(x)=\inf\{\Phi^*(\ell);\ell\in\mathcal{L}_0, T\ell+x\in C_0\}, x\in\mathcal{X}_0$$

We assume (H_{T1}): $T^T\mathcal{Y}_0\subset\mathcal{U}_0$, so that the duality schema is

$$\begin{array}{ccc} \Big\langle\ \mathcal{U}_0 & , & \mathcal{L}_0\ \Big\rangle \\ T^T\Big\uparrow & & \Big\downarrow T \\ \Big\langle\ \mathcal{Y}_0 & , & \mathcal{X}_0\ \Big\rangle \end{array} \qquad \text{(Schema 0)}$$

The analogue of F_0 for the dual problem is

$$G_0(y,u)\triangleq\inf_{\ell,x}\{\langle x,y\rangle-\langle \ell,u\rangle+F_0(\ell,x)\}=\inf_{x\in C_0}\langle x,y\rangle-\Phi_0^{**}(T^Ty+u)$$

where Φ_0^{**} is the $\sigma(\mathcal{U}_0,\mathcal{L}_0)$-lsc regularization of Φ. The corresponding value function is

$$\gamma_0(u)=\sup_{y\in\mathcal{Y}_0}\{\inf_{x\in C_0}\langle x,y\rangle-\Phi_0^{**}(T^Ty+u)\}, u\in\mathcal{U}_0$$

and the dual problem is (D_0).
Since C_0 is convex, F_0 is jointly convex. Assuming that C_0 is $\sigma(\mathcal{X}_0,\mathcal{Y}_0)$-closed, one obtains that for all $\ell\in\mathcal{L}_0$, the function $F_0(\ell,\cdot)$ is $\sigma(\mathcal{X}_0,\mathcal{Y}_0)$-closed. Therefore, assuming that C_0 is a $\sigma(\mathcal{X}_0,\mathcal{Y}_0)$-closed convex set, one can apply the general theory of Section 5 since the perturbation function F_0 satisfies the assumptions (29) and (31). In particular, we have by Theorem 5.2-(c): $\inf(P_0)=-\mathrm{cl}\,(-\gamma_0)(0)$.

Proposition 6.1. *Let us assume* (H_Φ) *and* (H_{T1}).
(a) *We have the little dual equality*

$$\inf\{\Phi^*(\ell);T\ell=x,\ell\in\mathcal{L}_0\}=\Lambda^*(x)\in[0,\infty], x\in\mathcal{X}_0. \tag{38}$$

(b) *If* C_0 *is convex and* $\sigma(\mathcal{X}_0,\mathcal{Y}_0)$*-closed, we have the dual equality*

$$\inf(P_0)=\sup(D_0)\in[0,\infty].$$

Proof. (a) is (b) with $C_0 = \{x\}$.
Let us show (b). Because of the identity $\inf(P_0) = -\mathrm{cl}\,(-\gamma_0)(0)$, by Theorem 5.2-(a) it remains to prove that $-\gamma_0$ is $\sigma(\mathcal{U}_0, \mathcal{L}_0)$-lsc at 0. Taking $y = 0$ instead of $\sup_{y\in\mathcal{Y}_0}$ in the definition of $\gamma_0(u)$, one obtains for all $u \in \mathcal{U}_0 : -\gamma_0(u) \le \Phi_0^{**}(u) \le \Phi(u)$. The norm $|\cdot|_\Phi$ is designed so that Φ is bounded above on a $|\cdot|_\Phi$-neighbourhood of zero. By the previous inequality, so is the convex function $-\gamma_0$. Therefore, $-\gamma_0$ is $|\cdot|_\Phi$-continuous on icordom $(-\gamma_0) \ni 0$. As it is convex and $\mathcal{L} = (\mathcal{U}_0, |\cdot|_\Phi)'$, it is also $\sigma(\mathcal{U}_0, \mathcal{L})$-lsc and a fortiori $\sigma(\mathcal{U}_0, \mathcal{L}_0)$-lsc, since $\mathcal{L} \subset \mathcal{L}_0$. $\square$

6.2. **An improvement.** We are going to consider the following duality schema, see Section 3.5:

$$\begin{array}{ccc} \Big\langle\ \mathcal{U}_1 & , & \mathcal{L}\ \Big\rangle \\ T_1^T \Big\uparrow & & \Big\downarrow T \\ \Big\langle\ \mathcal{Y}_1 & , & \mathcal{X}\ \Big\rangle \end{array} \qquad \text{(Schema 1)}$$

which is associated with the optimization problems (P) and (D_1). We need some preliminary results, gathered in the next lemma whose proof is postponed to Section 6.7.

Lemma 6.2. *Let us assume* (H_Φ) *and* (H_T).
(a) $|\cdot|_\Phi$ *and* $|\cdot|_\Lambda$ *are norms*
(b) $\mathrm{dom}\,\Phi^* \subset \mathcal{L}$ *and* $\mathrm{dom}\,\Lambda^* \subset \mathcal{X}$
(c) $T(\mathrm{dom}\,\Phi^*) \subset \mathrm{dom}\,\Lambda^*$ *and* $T\mathcal{L} \subset \mathcal{X}$
(d) $T_1^T\mathcal{Y}_1 \subset \mathcal{U}_1$
(e) $\Lambda_0(y) = \Lambda_1(y)$ *for all* $y \in \mathcal{Y}_0$

Note that the inclusions $T\mathcal{L} \subset \mathcal{X}$ and $T_1^T\mathcal{Y}_1 \subset \mathcal{U}_1$ stated in this lemma are necessary to validate the above duality schema.
Let F_1, G_1 and γ_1 be the analogous functions to F_0, G_0 and γ_0. We obtain

$$\begin{aligned} F_1(\ell, x) &= \Phi^*(\ell) + \delta_C(T\ell + x), \ell \in \mathcal{L}, x \in \mathcal{X} \\ G_1(y, u) &= \inf_{x\in C}\langle x, y\rangle - \Phi_1^{**}(T_1^T y + u), y \in \mathcal{Y}_1, u \in \mathcal{U}_1 \end{aligned}$$

and

$$\gamma_1(u) = \sup_{y\in\mathcal{Y}_1}\{\inf_{x\in C}\langle x, y\rangle - \Phi_1^{**}(T_1^T y + u)\}, u \in \mathcal{U}_1$$

Proposition 6.3. *Assuming* (H_Φ), (H_T) *and* (H_C), *we have the dual equality*

$$\inf(P) = \sup(D_1) \in [0, \infty]. \tag{39}$$

Proof. Because of (H_C), F_1 is jointly convex and $F_1(\ell, \cdot)$ is $\sigma(\mathcal{X}, \mathcal{Y}_1)$-closed convex for all $\ell \in \mathcal{L}$. As $T_1^T\mathcal{Y}_1 \subset \mathcal{U}_1$ (Lemma 6.2), one can apply the approach of Section 5 to the duality schema 1. Therefore, by Theorem 5.2-(c), the dual equality holds if γ_1 is $\sigma(\mathcal{U}_1, \mathcal{L})$-lsc at 0. As for Schema 0, we have $-\gamma_1(u) \le \Phi_1^{**}(u)$, for all $u \in \mathcal{U}_1$. Φ_1^{**} is the $\sigma(\mathcal{U}_1, \mathcal{L})$-lsc regularization of $\tilde\Phi(u) = \begin{cases} \Phi(u) & \text{if } u \in \mathcal{U}_0 \\ +\infty & \text{otherwise} \end{cases}$ and Φ is bounded above by 1 on the ball $\{u \in \mathcal{U}_0; |u|_\Phi < 1\}$. As $\mathcal{L} = (\mathcal{U}_1, |\cdot|_\Phi)'$, Φ_1^{**} is also the $|\cdot|_\Phi$-regularization of $\tilde\Phi$. Therefore,

Φ_1^{**} is bounded above by 1 on $\{u \in \mathcal{U}_1; |u|_\Phi < 1\}$, since $\{u \in \mathcal{U}_0; |u|_\Phi < 1\}$ is $|\cdot|_\Phi$-dense in $\{u \in \mathcal{U}_1; |u|_\Phi < 1\}$. As $-\gamma_1(\leq \Phi_1^{**})$ is convex and bounded above on a $|\cdot|_\Phi$-neighbourhood of 0, it is $|\cdot|_\Phi$-continuous on icordom $(-\gamma_1) \ni 0$. Hence, it is $\sigma(\mathcal{U}_1, \mathcal{L})$-lsc at 0. This completes the proof of the proposition. □

Corollary 6.4. *Assuming* (H_Φ) *and* (H_T)*, we have*

(a) (P_0) *and* (P) *are equivalent: they have the same solutions and* $\inf(P_0) = \inf(P) \in [0, \infty]$.
(b) $\Lambda_1^*(x) = \Lambda^*(x)$*, for all* $x \in \mathcal{X}$.

Proof. (a) is a direct consequence of dom $\Phi^* \subset \mathcal{L}$ and $T\mathcal{L} \subset \mathcal{X}$, see Lemma 6.2.
Let us show (b). By Proposition 6.3, for all $x \in \mathcal{X}$, we have $\inf\{\Phi^*(\ell); T\ell = x, \ell \in \mathcal{L}\} = \Lambda_1^*(x)$. But dom $\Phi^* \subset \mathcal{L}$ so that $\inf\{\Phi^*(\ell); T\ell = x, \ell \in \mathcal{L}\} = \inf\{\Phi^*(\ell); T\ell = x, \ell \in \mathcal{L}_0\}$. By (38): $\inf\{\Phi^*(\ell); T\ell = x, \ell \in \mathcal{L}_0\} = \Lambda^*(x)$. One concludes bringing these three identities together. □

6.3. **Primal attainment.** One proves the existence of solutions to (P) showing that Φ^* is inf-compact. As dom $\Phi^* \subset \mathcal{L}$, the restriction of Φ^* to $\mathcal{L}$ is also denoted by Φ^*.

Proposition 6.5. *Let us assume* (H_Φ), (H_T) *and* (H_C)*. We have*

(a) $\inf(P_0) = \inf_{x \in C_0} \Lambda^*(x) = \inf_{x \in C} \Lambda^*(x) \in [0, \infty]$
(b) *If in addition* $\inf(P_0) < \infty$*, then* (P_0) *is attained in* $\mathcal{L}$.
(c) *Let* $\bar{\ell} \in \mathcal{L}$ *be a solution to* (P_0)*, then* $\bar{x} \triangleq T\bar{\ell}$ *is a solution to* $(P_0^{\mathcal{X}})$ *and* $\inf(P_0) = \Phi^*(\bar{\ell}) = \Lambda^*(\bar{x})$.

The following lemma is needed for the proof of the proposition.

Lemma 6.6. *Under the hypothesis* (H_Φ)*,* Φ^* *is* $\sigma(\mathcal{L}, \mathcal{U}_1)$*-inf-compact.*

Proof of the lemma. For all $\ell \in \mathcal{L}_0$ and $\alpha > 0$, Young's inequality yields: $\langle \ell, u \rangle = \alpha \langle \ell, u/\alpha \rangle \leq [\Phi(u/\alpha) + \Phi^*(\ell)]\alpha$, for all $u \in \mathcal{U}_0$. Hence, for any $\alpha > |u|_\Phi$, $\langle \ell, u \rangle \leq [1 + \Phi^*(\ell)]\alpha$. It follows that $\langle \ell, u \rangle \leq [1 + \Phi^*(\ell)]|u|_\Phi$. Considering $-u$ instead of u, one gets

$$|\langle \ell, u \rangle| \leq [1 + \Phi^*(\ell)]|u|_\Phi, \forall u \in \mathcal{U}_0, \ell \in \mathcal{L}_0. \tag{40}$$

By completion, one deduces that for all $\ell \in \mathcal{L}$ and $u \in \mathcal{U}_1$, $|\langle \ell, u \rangle| \leq [1 + \Phi^*(\ell)]|u|_\Phi$. Hence, $\Phi^*(\ell) \leq A$ implies that $|\ell|_\Phi^* \leq A + 1$ where $|\ell|_\Phi^*$ stands for the uniform dual norm of $\ell \in \mathcal{L} = \mathcal{U}_1'$. Therefore, the level set $\{\Phi^* \leq A\}$ is relatively $\sigma(\mathcal{L}, \mathcal{U}_1)$-compact.
By construction, Φ^* is $\sigma(\mathcal{L}_0, \mathcal{U}_0)$-lsc and a fortiori $\sigma(\mathcal{L}, \mathcal{U}_1)$-lsc. Hence, $\{\Phi^* \leq A\}$ is $\sigma(\mathcal{L}, \mathcal{U}_1)$-closed and $\sigma(\mathcal{L}, \mathcal{U}_1)$-compact. □

Proof of the Proposition 6.5. Let us begin proving (b). By Lemma 6.2, T maps $\mathcal{L}$ into $\mathcal{X}$ and $T^T\mathcal{Y}_1 \subset \mathcal{U}_1$. It follows that T is $\sigma(\mathcal{L}, \mathcal{U}_1)$-$\sigma(\mathcal{X}, \mathcal{Y}_1)$-continuous. Indeed, for all $y \in \mathcal{Y}_1$, $\ell \mapsto \langle T\ell, y \rangle_{\mathcal{X}, \mathcal{Y}_1} = \langle \ell, T^T y \rangle_{\mathcal{L}, \mathcal{U}_1}$ is $\sigma(\mathcal{L}, \mathcal{U}_1)$-continuous. Since C is $\sigma(\mathcal{X}, \mathcal{Y}_1)$-closed, $\{\ell \in \mathcal{L}; T\ell \in C\}$ is $\sigma(\mathcal{L}, \mathcal{U}_1)$-closed.
As Φ^* is $\sigma(\mathcal{L}, \mathcal{U}_1)$-inf-compact (Lemma 6.6), it achieves its infimum on the closed set $\{\ell \in \mathcal{L}; T\ell \in C\}$ if $\inf(P) = \inf(P_0) < \infty$.
Let us show (a). The dual equality (39) gives us, for all $x_o \in C$, $\inf(P) = \sup_{y \in \mathcal{Y}_1}\{\inf_{x \in C}\langle x, y \rangle -$

$\Lambda_1(y)\} \leq \sup_{y\in\mathcal{Y}_1}\{\langle x_o, y\rangle - \Lambda_1(y)\} = \Lambda_1^*(x_o) = \Lambda^*(x_o)$ where the last equality comes from Corollary 6.4. Therefore

$$\inf(P) \leq \inf_{x\in C} \Lambda^*(x). \tag{41}$$

In particular, equality holds instead of inequality if $\inf(P) = +\infty$. Suppose now that $\inf(P) < \infty$. From statement (b), we already know that there exists $\bar{\ell} \in \mathcal{L}$ such that $\bar{x} \overset{\Delta}{=} T\bar{\ell} \in C$ and $\inf(P) = \Phi^*(\bar{\ell})$. Clearly $\inf(P) \leq \inf\{\Phi^*(\ell); T\ell = \bar{x}, \ell \in \mathcal{L}\} \leq \Phi^*(\bar{\ell})$. Hence, $\inf(P) = \inf\{\Phi^*(\ell); T\ell = \bar{x}, \ell \in \mathcal{L}\}$. By the little dual equality (38) we have $\inf\{\Phi^*(\ell); T\ell = \bar{x}, \ell \in \mathcal{L}\} = \Lambda^*(\bar{x})$. Finally, we have obtained $\inf(P) = \Lambda^*(\bar{x})$ with $\bar{x} \in C$. Together with (41), this leads us to the desired identity: $\inf(P) = \inf_{x\in C} \Lambda^*(x)$. By Lemma 6.2: $\operatorname{dom} \Lambda^* \subset \mathcal{X}$, so that we also have $\inf_{x\in C_0} \Lambda^*(x) = \inf_{x\in C} \Lambda^*(x)$.
Finally, (c) is a by-product of the proof of (a). □

6.4. **Dual attainment.** We now consider the following duality schema

$$\begin{array}{ccc} \Big\langle\ \mathcal{L} & , & \mathcal{U}_2\ \Big\rangle \\ T\Big\downarrow & & \Big\uparrow T^* \\ \Big\langle\ \mathcal{X} & , & \mathcal{Y}_2\ \Big\rangle \end{array} \tag{Schema 2}$$

where $\mathcal{U}_2 = \mathcal{L}^\sharp$ and $\mathcal{Y}_2 = \mathcal{X}^\sharp$. The topologies are the respective weak topologies. The associated perturbation functions are

$$\begin{aligned} F_2(\ell, x) &= F_1(\ell, x) = \Phi^*(\ell) + \delta_C(T\ell + x), \ell \in \mathcal{L}, x \in \mathcal{X} \\ G_2(u, y) &= \inf_{x\in\mathcal{X}} \langle x, y\rangle - \bar{\Phi}(T^*y + u), u \in \mathcal{U}_2, y \in \mathcal{Y}_2 \end{aligned}$$

As $F_2 = F_1$, the primal problem is (P) and its value function is

$$\varphi(x) = \inf\{\Phi^*(\ell); T\ell + x \in C, \ell \in \mathcal{L}\}, x \in \mathcal{X}.$$

The dual problem is

$$\textit{maximize } \inf_{x\in\mathcal{X}} \langle x, y\rangle - \bar{\Phi}(T^*y), y \in \mathcal{Y}_2$$

In fact this problem is nothing else than (D_2). Indeed, for all $y \in \mathcal{Y}_2$, $\bar{\Phi}(T^*y) = \sup_{\ell\in\mathcal{L}}\{\langle T\ell, y\rangle_{\mathcal{X},\mathcal{Y}_2} - \Phi^*(\ell)\} = \sup_{x\in\mathcal{X}}\{\langle x, y\rangle - \inf_{\ell; T\ell = x} \Phi^*(\ell)\} = \sup_{x\in\mathcal{X}} \langle x, y\rangle - \Lambda^*(x)$ where the last identity is the little dual equality (38). This means that $\bar{\Phi}(T^*y) = \bar{\Lambda}(y)$ and this dual problem is (D_2).

We assume that the hypotheses (H_Φ), (H_T) and (H_C) are satisfied. In particular, as $F_2 = F_1$, one can apply the approach of Section 5 to the duality Schema 2.

Let us denote φ_1^{**} the $\sigma(\mathcal{X}, \mathcal{Y}_1)$-lsc regularization of φ and φ_2^{**} its $\sigma(\mathcal{X}, \mathcal{Y}_2)$-lsc regularization. Since $\mathcal{X}$ separates $\mathcal{Y}_1$, the inclusion $\mathcal{Y}_1 \subset \mathcal{Y}_2$ holds. It follows that $\varphi_1^{**}(0) \leq \varphi_2^{**}(0) \leq \varphi(0)$. But we have (39) which is $\varphi_1^{**}(0) = \varphi(0)$. Therefore, one also obtains $\varphi_2^{**}(0) = \varphi(0)$ which is the dual equality

$$\inf(P) = \sup(D_2), \tag{42}$$

and one can apply the general theory of Section 5. Theorem 5.2-(b) gives

$$\operatorname{argmax}(D_2) = -\partial\varphi(0) = \{\bar{y} \in \mathcal{Y}_2; \varphi(x) - \varphi(0) \geq \langle -\bar{y}, x\rangle, \forall x \in \mathcal{X}\} \tag{43}$$

On the other hand, $\varphi(x) = \inf(P_x)$ where (P_x) is the minimization problem of $\Phi^*(\ell)$ subject to $T\ell \in C - x$. As $C - x$ shares the same requirements as C one can apply Proposition 6.5 to (P_x). This leads us to $\varphi(x) = \inf_{x' \in C-x} \Lambda^*(x') \le \Lambda^*(x_o - x)$, for all $x_o \in C$. Under the constraint qualification (15), one can pick x_o in icordom Λ^* and the previous inequality together with the geometric version of Hahn-Banach theorem provide us with $\partial\varphi(0) \neq \emptyset$. Taking (43) into account, this implies that (D_2) is attained:

$$\operatorname{argmax}(D_2) \neq \emptyset \tag{44}$$

Let us prove Theorem 3.3-(b). Let us first note that T^* extends T^T. Indeed, as $\mathcal{X} \subset \mathcal{X}_0$, for all $\ell \in \mathcal{L}$ and $y \in \mathcal{Y}_0$, $\langle T^T y, \ell\rangle_{\mathcal{U}_0,\mathcal{L}_0} = \langle y, T\ell\rangle_{\mathcal{Y}_0,\mathcal{X}} = \langle T^* y, \ell\rangle_{\mathcal{U}_2,\mathcal{L}}$.
As $T\mathcal{L} \subset \mathcal{X}$, T^* is $\sigma(\mathcal{Y}_2,\mathcal{X})$-$\sigma(\mathcal{U}_2,\mathcal{L})$-continuous.
As a direct consequence of (16) we have $\bar\Lambda(\bar y) < \infty$ which implies Theorem 3.3-(a): $\bar y$ stands in the $\sigma(\mathcal{Y}_2,\mathcal{X})$-closure of dom Λ since $\bar\Lambda$ is the convex $\sigma(\mathcal{Y}_2,\mathcal{X})$-lsc extension of Λ.
It follows by the continuity of T^* that $T^*\bar y$ stands in the $\sigma(\mathcal{U}_2,\mathcal{L})$-closure of T^*dom Λ. But T^*dom $\Lambda = T^T$dom Λ, as T^* extends T^T. This proves Theorem 3.3-(b).

Let us prove Theorem 3.3-(c). Let $\bar y \in \operatorname{argmax}(D_2)$. By (43), for all $x \in \mathcal{X}$ and any $x_o \in C \cap$ icordom Λ^*, $\langle -\bar y, x\rangle \le \varphi(x) - \varphi(0) \le \Lambda^*(x_o - x) - \varphi(0) \le \Lambda^*(x_o - x)$. It follows that $\langle \bar y, x\rangle \le \Lambda^*(x_o) + 1$ for all $x \in D_{x_o} \triangleq \{x \in \mathcal{X}; \Lambda^*(x_o + x) \le \Lambda^*(x_o) + 1\}$. This implies that for all $x \in \mathcal{X}$, $\langle \bar y, x\rangle \le [1 + \Lambda^*(x_o)] j_{D_{x_o}}(x)$. Since $j_D(-x) = j_{-D}(x)$, we finally obtain

$$-[1 + \Lambda^*(x_o)] j_{-D_{x_o}}(x) \le \langle \bar y, x\rangle \le [1 + \Lambda^*(x_o)] j_{D_{x_o}}(x), \forall x \in \mathcal{X} \tag{45}$$

for any x_o standing in $C \cap$ icordom Λ^*.

6.5. **Dual representation of the minimizers.** We keep the framework of Schema 2 and derive the Kuhn-Tucker conditions in this situation. The Lagrangian associated with $F_2 = F_1$ and Schema 2 is for any $\ell \in \mathcal{L}, y \in \mathcal{Y}_2$,

$$\begin{aligned} K_2(\ell, y) &\triangleq \inf_{x \in \mathcal{X}} \{\langle x, y\rangle + \Phi^*(\ell) + \delta_C(T\ell + x)\}, \\ &= \Phi^*(\ell) - \langle T\ell, y\rangle + \inf_{x \in C} \langle x, y\rangle. \end{aligned}$$

Under the constraint qualification (15), the dual equality (42) holds and we have both the primal and dual attainments. Hence the Kuhn-Tucker conditions (Theorem 5.3) are satisfied. Let $(\bar\ell, \bar y) \in \mathcal{L} \times \mathcal{Y}_2$ be a solution to (P) and (D_2). It is a saddle-point of K_2 and the Kuhn-Tucker conditions (36) and (37) are $\partial_\ell K_2(\bar\ell, \bar y) \ni 0$ and $\partial_y(-K_2)(\bar\ell, \bar y) \ni 0$. Since $-\langle T\ell, y\rangle$ is locally weakly upper bounded as a function of y around $\bar y$ and as a function of ℓ around $\bar\ell$, one can apply ([27], Theorem 20) to derive $\partial_\ell K_2(\bar\ell, \bar y) = \partial\Phi^*(\bar\ell) - T^*\bar y$ and $\partial_y(-K_2)(\bar\ell, \bar y) = \partial(-\inf_{x \in C}\langle x, \cdot\rangle) + T\bar\ell$. Therefore the Kuhn-Tucker conditions are

$$\begin{aligned} T^*\bar y &\in \partial\Phi^*(\bar\ell) \\ -T\bar\ell &\in \partial(\delta^*_{-C})(\bar y) \end{aligned}$$

where δ^*_{-C} is the convex conjugate of the convex indicator of $-C$. Indeed, $-\inf_{x \in C}\langle x, y\rangle = \sup_{x \in -C}\langle x, y\rangle = \sup_{x \in \mathcal{X}}\{\langle x, y\rangle - \delta_{-C}(x)\} = \delta^*_{-C}(y)$. Both Φ^* and δ^*_{-C} are lsc functions as convex conjugates. Hence, considering the dual Schema 2, the following conjugate relations hold:

$$\begin{aligned} \bar{\ell} &\in \partial\bar{\Phi}(T^*\bar{y}) && (46)\\ \bar{y} &\in \partial\delta_{-\bar{C}}(-\bar{x}) && (47) \end{aligned}$$

where $\bar{x} \triangleq T\bar{\ell}$ and $\bar{C}$ stands for the $\sigma(\mathcal{X},\mathcal{Y}_2)$-closure of C. Of course, as C is $\sigma(\mathcal{X},\mathcal{Y}_1)$-closed by hypothesis (H_C), it is a fortiori $\sigma(\mathcal{X},\mathcal{Y}_2)$-closed, so that $\bar{C} = C$. Therefore, one can replace $\bar{C}$ by C in (47). The statements (46) and (47) are equivalent to the Young's identities

$$\begin{aligned} \Phi^*(\bar{\ell}) + \bar{\Phi}(T^*\bar{y}) &= \langle \bar{x}, \bar{y}\rangle && (48)\\ \delta_C(\bar{x}) + \delta^*_{-C}(\bar{y}) &= \langle -\bar{x}, \bar{y}\rangle && (49) \end{aligned}$$

It follows from (49) that $\delta_C(\bar{x}) < \infty$ which is equivalent to

$$\bar{x} \in C. \tag{50}$$

Now (49) is $-\langle \bar{x}, \bar{y}\rangle = \delta^*_{-C}(\bar{y}) = -\inf_{x\in C}\langle x, \bar{y}\rangle$:

$$\langle \bar{x}, \bar{y}\rangle = \inf_{x\in C}\langle x, \bar{y}\rangle \tag{51}$$

6.6. **Collecting the results.** The results are collected as follows.
Theorem 3.2: (a) is Proposition 6.1-(a); *(b)* is Corollary 6.4-(a); *(c)* is Propositions 6.3 and 6.5-(a); *(d)* is Proposition 6.5-(b) and a consequence of Lemma 6.6; *(e)* is Proposition 6.1-(b).
Theorem 3.3: The dual equality is (42); the dual attainment is (44); *(a)*, *(b)* and *(c)* have been proved at the end of Section 6.4.
Theorem 3.4: The first part is already proved. For the second part, *(a)* is (50), *(b)* is (51), *(c)* is (46), the first equality of (16) is (48), $\bar{x}$ minimizes Λ^* on C_0 by Proposition 6.5-(c), the representation of $\bar{y}$ together with the second equality of (16) follow from Corollary 3.5.
Corollary 3.5 is a direct application of the previous results.

Notice that the results of Corollary 3.5 which are used in the proof of the last statements of Theorem 3.4 are consequences of the first statements of this theorem.

6.7. **Proof of Lemma 6.2.**

Proof of Lemma 6.2. (a) Because of (H_Φ), $\{u \in \mathcal{U}_0; \max(\Phi(u), \Phi(-u)) \le 1\}$ is a convex, absorbing and balanced set. Hence, $|\cdot|_\Phi$ is a seminorm on $\mathcal{U}_0$. It is a norm under the assumption $(H_{\Phi 3})$. Indeed, let $u \in \mathcal{U}_0$ be such that $|u|_\Phi = 0$. It implies that for all real t, $0 \le \Phi(tu) \le 1$ and since Φ is convex $\Phi(tu) = \Phi(0) = 0$. By $(H_{\Phi 3})$, this implies that $u = 0$.
Because of (H_Φ) and (H_{T1}), $\{y \in \mathcal{Y}_0; \max(\Lambda(y), \Lambda(-y)) \le 1\}$ is also a convex, absorbing and balanced set. Hence, $|\cdot|_\Lambda$ is a seminorm on $\mathcal{Y}_0$. It is a norm under the assumption $(H_{\Phi 3})$ and (H_{T2}). Indeed, by $(H_{\Phi 3})$ as above, $|y|_\Lambda = 0$ implies that $T^T y = 0$ and by (H_{T2}): $y = 0$.
(b) It follows from (40) that dom $\Phi^* \subset \mathcal{L}$. One proves dom $\Lambda^* \subset \mathcal{X}$ similarly.
(c) Let us denote $\Phi_{max}(u) \triangleq \max(\Phi(u), \Phi(-u))$ and $\Lambda_{max}(y) \triangleq \max(\Lambda(y), \Lambda(-y))$, so that $|u|_\Phi = \inf\{\alpha > 0; \Phi_{max}(u/\alpha) \le 1\}$ and $|y|_\Lambda = \inf\{\alpha > 0; \Lambda_{max}(y/\alpha) \le 1\}$. Let us introduce the dual uniform norms $|\ell|^*_\Phi \triangleq \sup_{u,|u|_\Phi\le 1}|\langle \ell, u\rangle|$, $\ell \in \mathcal{L}$ and $|x|^*_\Lambda \triangleq$

$\sup_{y,|y|_\Lambda \le 1} |\langle x, y\rangle|$, $x \in \mathcal{X}$. Let us also consider $|\cdot|_{\Phi^*_{max}}$ and $|\cdot|_{\Lambda^*_{max}}$ the gauge functionals of the level sets $\{\Phi^*_{max} \le 1\}$ and $\{\Lambda^*_{max} \le 1\}$.
The little dual equality (38) for Φ_{max} and Λ_{max} implies that

$$\Lambda^*_{max}(x) \le \Phi^*_{max}(\ell), \text{ for all } \ell \in \mathcal{L}_0 \text{ and } x \in \mathcal{X}_0 \text{ such that } T\ell = x \tag{52}$$

Therefore, $T(\text{dom } \Phi^*_{max}) \subset \text{dom } \Lambda^*_{max}$. On the other hand, by Proposition 6.7 below, the linear space spanned by dom Φ^*_{max} is dom $|\cdot|_{\Phi^*_{max}}$ and the linear space spanned by dom Λ^*_{max} is dom $|\cdot|_{\Lambda^*_{max}}$. But, dom $|\cdot|_{\Phi^*_{max}} = \text{dom } |\cdot|^*_\Phi = \mathcal{L}$ and dom $|\cdot|_{\Lambda^*_{max}} = \text{dom } |\cdot|^*_\Lambda = \mathcal{X}$ by Proposition 6.7. Hence, $T\mathcal{L} \subset \mathcal{X}$.
(d) Let us show that $T : (\mathcal{L}, |\cdot|^*_\Phi) \to (\mathcal{X}, |\cdot|^*_\Lambda)$ is continuous. We know by Proposition 6.7 that $|\cdot|_{\Phi^*_{max}}$ and $|\cdot|^*_\Phi$ as well as $|\cdot|_{\Lambda^*_{max}}$ and $|\cdot|^*_\Lambda$ are equivalent norms on $\mathcal{L}$ and $\mathcal{X}$ respectively. For all $\ell \in \mathcal{L}$, $|T\ell|^*_\Lambda \le 2|T\ell|_{\Lambda^*_{max}} = 2\inf\{\alpha > 0; \Lambda^*_{max}(T\ell/\alpha) \le 1\} \le 2\inf\{\alpha > 0; \Phi^*_{max}(\ell/\alpha) \le 1\}$. This last inequality follows from (52). Going on, we get $|T\ell|^*_\Lambda \le 2|\ell|_{\Phi^*_{max}} \le 4|\ell|^*_\Phi$, which proves that T shares the desired continuity property.
Let us take $y \in \mathcal{Y}_1$. For all $\ell \in \mathcal{L}$, $|\langle T_1^T y, \ell\rangle_{\mathcal{L}^\sharp,\mathcal{L}}| = |\langle y, T\ell\rangle_{\mathcal{Y}_1,\mathcal{X}}| \le |y|_\Lambda |T\ell|^*_\Lambda \le 4|y|_\Lambda |\ell|^*_\Phi$. Hence, $T_1^T y$ stands in the topological bidual space of $(\mathcal{U}_0, |\cdot|_\Phi)$. More, it is the strong limit of a sequence in $\mathcal{U}_0$. Indeed, there exists a sequence (y_n) in $\mathcal{Y}_0$ such that $\lim_{n\to\infty} y_n = y$ in $(\mathcal{Y}_1, |\cdot|_\Lambda)$. Hence, for all $\ell \in \mathcal{L}$, $|\langle T_1^T y_n - T_1^T y, \ell\rangle_{\mathcal{L}^\sharp,\mathcal{L}}| = |\langle y_n - y, T\ell\rangle_{\mathcal{Y}_1,\mathcal{X}}| \le 4|y_n - y|_\Lambda |\ell|^*_\Phi$ and $\sup_{\ell\in\mathcal{L},|\ell|^*_\Phi \le 1} |\langle T_1^T y_n - T_1^T y, \ell\rangle| \le 4|y_n - y|_\Lambda$ tends to 0 as n tends to infinity, where $T_1^T y_n = T^T y_n$ belongs to $\mathcal{U}_0$ for all $n \ge 1$. Consequently, $T_1^T y \in \mathcal{U}_1$.
(e) For all $y \in \mathcal{Y}_0$, $\Lambda_0(y) \triangleq \Phi_0^{**}(T^T y) = \sup_{\ell\in\mathcal{L}_0}\{\langle T^T y, \ell\rangle - \Phi^*(\ell)\} = \sup_{\ell\in\mathcal{L}}\{\langle T^T y, \ell\rangle - \Phi^*(\ell)\} = \Phi_1^{**}(T^T y) \triangleq \Lambda_1(y)$, where the identity $\sup_{\ell\in\mathcal{L}_0} = \sup_{\ell\in\mathcal{L}}$ comes from dom $\Phi^* \subset \mathcal{L}$. □

6.8. **Gauge functionals associated with a convex function.** Let $\theta : S \to [0,\infty]$ be an extended nonnegative convex function on a vector space S, such that $\theta(0) = 0$. Let $S^\sharp$ be the algebraic dual space of S and θ^* the convex conjugate of θ :

$$\theta^*(r) \triangleq \sup_{s\in S}\{\langle r, s\rangle - \theta(s)\}, r \in S^\sharp.$$

It is easy to show that $\theta^* : S^\sharp \to [0,\infty]$ and $\theta^*(0) = 0$. We denote $C_\theta \triangleq \{\theta \le 1\}$ and $C_{\theta^*} \triangleq \{\theta^* \le 1\}$ the unit level sets of θ and θ^*. The gauge functionals to be considered are

$$\begin{aligned} j_\theta(s) &\triangleq \inf\{\alpha > 0; \alpha s \in C_\theta\} = \inf\{\alpha > 0; \theta(s/\alpha) \le 1\} \in [0,\infty], s \in S. \\ j_{\theta^*}(r) &\triangleq \inf\{\alpha > 0; \alpha r \in C_{\theta^*}\} = \inf\{\alpha > 0; \theta^*(r/\alpha) \le 1\} \in [0,\infty], r \in S^\sharp. \end{aligned}$$

As 0 belongs to C_θ and C_{θ^*}, one easily proves that j_θ and j_{θ^*} are positively homogeneous. Similarly, as C_θ and C_{θ^*} are convex sets, j_θ and j_{θ^*} are convex functions.

Proposition 6.7. *Let $\theta : S \to [0,\infty]$ be an extended nonnegative convex function on a vector space S, such that $\theta(0) = 0$ as above, then for all $r \in S^\sharp$, we have*

$$\frac{1}{2} j_{\theta^*}(r) \le \delta^*_{C_\theta}(r) \triangleq \sup_{s\in C_\theta}\langle r, s\rangle \le 2 j_{\theta^*}(r).$$

We also have

$$\text{cone dom } \theta^* = \text{dom } j_{\theta^*} = \text{dom } \delta^*_{C_\theta}$$

where cone dom θ^* *is convex cone (with vertex 0) generated by* dom θ^*.

Proof. • Let us first show that $\delta^*_{C_\theta}(r) \le 2j_{\theta^*}(r)$ for all $r \in S^\sharp$. If $j_{\theta^*}(r) > 0$, then for all $s \in C_\theta$, $\langle r, s\rangle = \langle r/j_{\theta^*}(r), s\rangle j_{\theta^*}(r) \le [\theta(s) + \theta^*(r/j_{\theta^*}(r))]j_{\theta^*}(r) \le (1+1)j_{\theta^*}(r)$.
If $j_{\theta^*}(r) = 0$, then $\theta^*(tr) \le 1$ for all $t > 0$. For any $s \in C_\theta$, we get $\langle r, s\rangle = \frac{1}{t}\langle tr, s\rangle \le \frac{1}{t}[\theta(s) + \theta^*(tr)] \le 2/t$. Letting t tend to infinity, one obtains that $\langle r, s\rangle \le 0$.
• Let us show that $j_{\theta^*}(r) \le 2\delta^*_{C_\theta}(r)$. If $\delta^*_{C_\theta}(r) = \infty$, there is nothing to prove. So, let us suppose that $\delta^*_{C_\theta}(r) < \infty$. As $0 \in C_\theta$, we have $\delta^*_{C_\theta}(r) \ge 0$.
First case: $\delta^*_{C_\theta}(r) > 0$. For all $s \in S$ and $\epsilon > 0$, we have $s/[j_\theta(s) + \epsilon] \in C_\theta$. It follows that $\langle r/\delta^*_{C_\theta}(r), s\rangle = \langle r, s/[j_\theta(s)+\epsilon]\rangle \frac{j_\theta(s)+\epsilon}{\delta^*_{C_\theta}(r)} \le \delta^*_{C_\theta}(r)\frac{j_\theta(s)+\epsilon}{\delta^*_{C_\theta}(r)} = j_\theta(s) + \epsilon$. Therefore, $\langle r/\delta^*_{C_\theta}(r), s\rangle \le j_\theta(s)$, for all $s \in S$.
If s doesn't belong to C_θ, then $j_\theta(s) \le \theta(s)$. This follows from the the assumptions on θ : convex function such that $\theta(0) = 0 = \min\theta$ and the positive homogeneity of j_θ. Otherwise, if s belongs to C_θ, we have $j_\theta(s) \le 1$. Hence, $\langle r/\delta^*_{C_\theta}(r), s\rangle \le \max(1, \theta(s)), \forall s \in S$. On the other hand, there exists $s_o \in S$ such that $\theta^*(r/[2\delta^*_{C_\theta}(r)]) \le \langle r/[2\delta^*_{C_\theta}(r)], s_o\rangle - \theta(s_o) + 1/2$. The last two inequalities provide us with $\theta^*(r/[2\delta^*_{C_\theta}(r)]) \le \frac{1}{2}\max(1, \theta(s_o)) - \theta(s_o) + \frac{1}{2} \le 1$ since $\theta(s_o) \le 0$. We have proved that $j_{\theta^*}(r) \le 2\delta^*_{C_\theta}(r)$.
Second case: $\delta^*_{C_\theta}(r) = 0$. We have $\langle r, s\rangle \le 0$ for all $s \in C_\theta$. As dom C_θ is a subset of the cone generated by C_θ, we also have for all $t > 0$ and $s \in$ dom θ, $\langle tr, s\rangle \le 0$. Hence $\langle tr, s\rangle - \theta(s) \le 0$ for all $s \in S$ and $\theta^*(tr) \le 0$, for all $t \ge 0$. As $\theta^* \ge 0$, we have $\theta^*(tr) = 0$, for all $t \ge 0$. It follows that $j_{\theta^*}(r) = 0$. This completes the proof of the equivalence of j_{θ^*} and $\delta^*_{C_\theta}$.
• Finally, this equivalence implies that dom j_{θ^*} = dom $\delta^*_{C_\theta}$ and as $\theta^*(0) = 0$ we have $0 \in$ dom θ^* which implies that cone dom θ^* = dom j_{θ^*}. □

7. The proof of Theorem 2.1

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

Proposition 7.1. *Under the assumptions of Theorem 2.1, for all closed subset F of $\mathcal{X}$, we have*

$$\limsup_{\delta\to 0}\limsup_{n\to\infty}\frac{1}{n}\log I\!P(X_n \in F \mid X_n \in C_\delta) \le -J_{C_0}(F)$$

where

$$J_{C_0}(x) \triangleq \begin{cases} J(x) - J(C_0) & \text{if } x \in C_0 \\ +\infty & \text{if } x \notin C_0 \end{cases}$$

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

$$\begin{aligned} & \limsup_n \frac{1}{n}\log I\!P(X_n \in F \mid X_n \in C_\delta) \\ \le\ & \limsup_n \frac{1}{n}\log I\!P(X_n \in F \cap \text{cl } C_\delta) - \liminf_n \frac{1}{n}\log I\!P(X_n \in \text{int } C_\delta) \\ \le\ & -J(F \cap \text{cl } C_\delta) + J(\text{int } C_\delta) \\ \le\ & -J(F \cap \text{cl } C_\delta) + J(C_0) \end{aligned}$$

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

Lemma 7.2. *For any closed set* F, $\lim_\delta J(F\cap \mathrm{cl}\ C_\delta) = J(F\cap C_0)\in [0,\infty]$.

Proof. As $C_0\subset \mathrm{cl}\ C_\delta$, for all $\delta>0$, we have $J(F\cap \mathrm{cl}\ C_\delta)\leq J(F\cap C_0)$. Since C_δ is nonincreasing, $J(F\cap \mathrm{cl}\ C_\delta)$ is nondecreasing and $\lim_\delta J(F\cap \mathrm{cl}\ C_\delta)=\sup_\delta J(F\cap \mathrm{cl}\ C_\delta)\in [0,\infty]$. If $\sup_\delta J(F\cap \mathrm{cl}\ C_\delta)=\infty$, the inequality $J(F\cap \mathrm{cl}\ C_\delta)\leq J(F\cap C_0)$ leads to the desired result.
Now, let us suppose that $\sup_\delta J(F\cap \mathrm{cl}\ C_\delta)<\infty$. As $F\cap \mathrm{cl}\ C_\delta$ is closed and J is inf-compact, for any δ there exists $x_\delta\in F\cap \mathrm{cl}\ C_\delta$ such that $J(x_\delta)=J(F\cap \mathrm{cl}\ C_\delta)$ and one can extract a converging subsequence $x_k\to x_*$. Because the C_δ's are nonincreasing, we get $\cap_\delta \mathrm{cl}\ C_\delta = \cap_k \mathrm{cl}\ C_{\delta_k}$ and $\lim_\delta J(F\cap \mathrm{cl}\ C_\delta)=\lim_k J(x_k)$. More, $x_*\in F\cap(\cap_k \mathrm{cl}\ C_{\delta_k})=F\cap C_0$ and as J is lsc: $\lim_k J(x_k)\geq J(x_*)\geq J(F\cap C_0)$. Therefore, $\lim_\delta J(F\cap \mathrm{cl}\ C_\delta)\geq J(F\cap C_0)$ which completes the proof. □

Let us state the lower bound corresponding to Proposition 7.1.

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

$$C_0\subset \mathrm{int}\ C_\delta, \forall\delta>0 \tag{53}$$

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

$$\liminf_{\delta\to 0}\liminf_{n\to\infty}\frac{1}{n}\log I\!P(X_n\in G\mid X_n\in C_\delta)\geq -J_{C_0}(G).$$

Proof. For all $\delta>0$,

$$\begin{aligned}
&\liminf_n \frac{1}{n}\log I\!P(X_n\in G\mid X_n\in C_\delta)\\
\geq\ &\liminf_n \frac{1}{n}\log I\!P(X_n\in G\cap \mathrm{int}\ C_\delta)-\limsup_n\frac{1}{n}\log I\!P(X_n\in \mathrm{cl}\ C_\delta)\\
\geq\ & -J(G\cap \mathrm{int}\ C_\delta)+J(\mathrm{cl}\ C_\delta)\\
\geq\ & -J(G\cap C_0)+J(\mathrm{cl}\ C_\delta).
\end{aligned}$$

One concludes with the above lemma. □

Let us recall that $\mathcal{H}=\mathrm{argmin} J_{C_o}$ is the set of the minimizers of J on C_0. As C_0 is closed and J is inf-compact, $\mathcal{H}$ is a compact set. As an immediate corollary of Proposition 7.1, we have the following conditional law of large numbers which is the part of the statement of Theorem 2.1 concerning X_n.

Proposition 7.4. *For all open subset* H *of* $\mathcal{X}_0$ *such that* $\mathcal{H}\subset H$, *we have*

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

Let us now have a look at L_n. We are interested in the asymptotic behavior of $I\!P(L_n\in \cdot\mid TL_n\in C_\delta)$ with $C_\delta\subset\mathcal{X}_0$. Let us denote $A_\delta\triangleq T^{-1}(C_\delta)=\{\ell\in\mathcal{L}_0; T\ell\in C_\delta\}$ and $A_0=T^{-1}C_0$. It is useful to state the assumptions on the C_δ's rather than on the A_δ's. In fact, one has the following transfer result.

Lemma 7.5. *We assume that* T *is continuous.*

(a) *If* C_0 *is closed and* $J(C_0) = J(\text{int } C_0)$, *then* A_0 *is closed and* $I(A_0) = I(\text{int } A_0)$.

(b) *If* $C_0 \triangleq \cap_\delta \text{cl } C_\delta \subset \text{int } C_\delta$ *for all* $\delta > 0$, *then* $A_0 = \cap_\delta \text{cl } A_\delta$ *and* $A_0 \subset \text{int } A_\delta$ *for all* $\delta > 0$.

Proof. Since T is continuous, for any $A = T^{-1}C$, we have: $T^{-1}(\text{int } C) \subset \text{int } A \subset A \subset \text{cl } A \subset T^{-1}(\text{cl } C)$.
Let us begin with (a). As C_0 is closed, so is A_0. For any $A = T^{-1}C$, we have $I(A) = \inf\{\Phi^*(\ell); T\ell \in C\} = \inf_{x\in C} \inf\{\Phi^*(\ell); T\ell = x\} = \inf_{x\in C} J(x) = J(C)$. Hence, $I(A_0) = J(C_0) = J(\text{int } C_0)$ (by hypothesis) $= I(T^{-1}(\text{int } C_0) \geq I(\text{int } A_0)$, since $T^{-1}(\text{int } C_0) \subset \text{int } A_0$. But the converse inequality: $I(A_0) \leq I(\text{int } A_0)$ is clear.
Let us prove (b). We have: $\cap_\delta \text{cl } A_\delta \subset T^{-1}(\cap_\delta \text{cl } C_\delta) \triangleq A_0 \triangleq T^{-1}(C_0) \subset T^{-1}(\cap_\delta \text{int } C_\delta)$ (by hypothesis) $\subset \cap_\delta \text{int } A_\delta \subset \cap_\delta \text{cl } A_\delta$. This proves that all these sets are equal, and in particular that $A_0 = \cap_\delta \text{cl } A_\delta$. On the other hand, as for any $\delta > 0$, $C_0 \subset \text{int } C_\delta$, we have $A_0 = T^{-1}(C_0) \subset T^{-1}(\text{int } C_\delta) \subset \text{int } A_\delta$. $\square$

Let us recall that $\mathcal{G}$ is the set of the minimizers of I on A_0. By the above Lemma 7.5, in the situation of the L_n's, Proposition 7.4 becomes the following

Proposition 7.6. *Under our general assumptions, for all open subset* G *of* $\mathcal{L}_0$ *such that* $\mathcal{G} \subset G$, *we have*

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

Note that by Lemma 7.5, the A_δ's share the same properties as the C_δ's. In particular, $I(\text{int } A_\delta) \leq I(A_0)$ also holds for them.

References

[1] J.M. Borwein, A.S. Lewis. *Duality relationships for entropy-like minimization problems.* SIAM J. Control and Optim. 29, 1991, 325-338.

[2] J.M. Borwein, A.S. Lewis. *On the convergence of moment problems.* Trans. Amer. Math. Soc. 325, 1991, 249-271.

[3] J.M. Borwein, A.S. Lewis. *Convergence of best entropy estimates.* SIAM J. Optim. 1, 1991, 191-205.

[4] J.M. Borwein, A.S. Lewis. *Decomposition of multivariate functions.* Can. J. Math. Vol. 44(3), 1992, 463-482.

[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, R.D. Nussbaum. *Entropy minimization, DAD problems and doubly stochastic kernels.* J. Funct. Anal. 123, 1994, 264-307.

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

[8] P. Cattiaux, C. Léonard. *Large deviations and Nelson's processes.* Forum Math. 7, 1995, 95-115.

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

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

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

[12] A.B. Cruzeiro, Liming Wu, J.C. Zambrini. Bernstein processes associated with a Markov process, in: *Proceedings of the Third International Workshop on Stochastic Analysis and Mathematical Physics* ANESTOC'98, Ed. R. Rebolledo. Trends in Mathematics, Birkhaüser.

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

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

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

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

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

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

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

[20] C. Léonard. *Dominating points and entropic projections.* 2003, preprint.

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

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

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

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

[25] S.T. Rachev, L. Rüschendorf. *Mass Transportation Problems*, Probability and its Applications, Springer (1998).

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

[27] R.T. Rockafellar. *Conjugate Duality and Optimization*, Regional Conferences Series in Applied Mathematics, SIAM, Philadelphia (1974).

[28] L. Rüschendorf, W. Thomsen. *Note on the Schrödinger equation and I-projections.* Stat. Probab. Letters, 17, 1993, 369-375.

[29] E. Schrödinger. *Sur la théorie relativiste de l'électron et l'interprétation de la mécanique quantique.* Ann. Inst. Henri Poincaré 2, 1932, 269-310.

[30] V. Strassen. *The existence of probability measures with given marginals.* Ann. Math. Stat. 36(2), 1965, 423-439.

[31] D.W. Stroock, O. Zeitouni. Microcanonical distributions, Gibbs states and the equivalence of ensembles. In: *Festchrift in Honour of F. Spitzer.* Eds. R. Durrett and H. Kesten, Birkhaüser (1991), 399-424.

[32] M. Teboulle, I. Vajda. *Convergence of best ϕ-entropy estimates.* IEEE Trans. Inform. Theory 39, 1993, 297-301.

[33] J.C. Zambrini. *Variational processes and stochastic versions of mechanics.* J. Math. Phys. 27, 1986, pp. 2307-.

Modal-X, Université Paris 10, Bât. G, 200 av. de la République. 92001 Nanterre Cedex France

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