Some Results about Entropic Projections

C. Léonard

To appear in the proceedings of the Conference on *Stochastic Analysis and Mathematical Physics*, Lisbon, 1999. Eds. A. B. Cruseiro and J. C. Zambrini.

1. Introduction

We give a short survey of some of our results related to the maximum entropy method. In this article, the Large Deviations approach is privileged, rather than the more direct convex analytical approach. Indeed, the proposed applications are naturally stated in terms of large random particle systems. These are the existence and construction problems for the Schrödinger's bridges and the Nelson's diffusion processes. These problems arise from probabilistic approaches to Quantum Mechanics.

Let R be a fixed reference probability measure. The optimization problem to be investigated is the minimization of the relative entropy $I(\cdot \mid R)$ subject to a general linear constraint ($-I(\cdot \mid R)$ is concave, it is *the* entropy to be maximized).
We introduce a constraint function φ which allows the description of a general infinite dimensional linear constraint. Our assumptions on φ are exponential integrability conditions with respect to R; they are called *Cramér's conditions*: the very strong one (2.5), the strong one (4.1) and the weak one (2.2).

At Section 2, the dual equality of our optimization problem is obtained under the very strong Cramér's condition. It yields a criterion of existence of a unique minimizer: the *I-projection* (see Definition 2.1).

At Section 3, this criterion is specified for the problems of existence of the Schrödinger's bridges and the Nelson's diffusion processes.

At Section 4, we give a characterization of the I-projection, under the strong Cramér's condition. This result definitely improves the previous related results in the literature.

At Section 5, we give a brief glimpse of the situation when φ is not very integrable: under the weak Cramér's condition. In this situation, it may happen that the I-projection doesn't exist anymore. The minimization problem has to be replaced by an "extended" one. Its minimizers may not be unique anymore and may admit a singular part which is not a measure (lack of σ-additivity). In order to keep our Large Deviations approach, we state an extension of Sanov's theorem at Theorem 5.1. The situation is illustrated with a simple example at Subsections and . We call *entropic projection* the minimizers of the extended relative entropy: the rate function of the extended Sanov's theorem (see Definition 5.3). It appears that I-projections are entropic projections.

Recall that a sequence $\{X_n\}_{n\geq 1}$ obeys the *large deviation principle* (LDP) in a topological space with the rate function I if for any measurable subset A :

$$\begin{aligned} -\inf_{x\in\mathrm{int}(A)} I(x) &\leq \liminf_{n\to\infty} \frac{1}{n} \log P(X_n \in A) \\ &\leq \limsup_{n\to\infty} \frac{1}{n} \log P(X_n \in A) \leq -\inf_{x\in\mathrm{cl}(A)} I(x) \end{aligned}$$

The rate function I is said to be a good rate function if it is inf-compact. For more details about large deviations, see the book of A. Dembo and O. Zeitouni [11].

2. Under the very strong Cramér's condition

2.1. Sanov's theorem

Let $(Z_i)_{i\geq 1}$ be an independent identically distributed sequence of random elements with values in a measurable space $(\Omega, \mathcal{A})$ and common law $R \in \mathcal{P}(\Omega)$: the set of probability measures on $(\Omega, \mathcal{A})$. Sanov's theorem ([11], Theorem 6.2.10) states that the sequence of empirical measures $L_n = \frac{1}{n}\sum_{i=1}^n \delta_{Z_i} \in \mathcal{P}(\Omega)$ (δ stands for Dirac measure) satisfies a LDP in $\mathcal{P}(\Omega)$ with the weak topology $\sigma(\mathcal{P}(\Omega), B)$, where B is the space of measurable bounded functions on Ω. Its good rate function is the relative entropy with respect to R, it is given for any $P \in \mathcal{P}(\Omega)$ by

$$\begin{aligned} I(P \mid R) &= \sup_{f\in B} \left\{ \int_\Omega f\, dR - \log \int_\Omega e^f\, dR \right\} \\ &= \begin{cases} \int_\Omega \log\left(\frac{dP}{dR}\right) dP & \text{if } P \ll R \\ +\infty & \text{otherwise} \end{cases} \end{aligned} \tag{2.1}$$

2.2. A constraint function

One considers a function $\varphi : \Omega \mapsto \mathcal{X}$ on Ω with its values in a vector space $\mathcal{X}$ in separating duality with a vector space $\mathcal{Y}$. The space $\mathcal{X}$ is endowed with the σ-field generated by the linear forms $x \in \mathcal{X} \mapsto \langle x, y\rangle \in \mathbb{R}$, $y \in \mathcal{Y}$. It is also assumed that φ is measurable in the following sense: $\omega \in \Omega \mapsto \langle y, \varphi(\omega)\rangle \in \mathbb{R}$ is measurable for all $y \in \mathcal{Y}$.

2.3. Cramér's theorem

The sequence $X_i = \varphi(Z_i), i \geq 1$ is independent identically distributed on $\mathcal{X}$ with common law $R \circ \varphi^{-1}$. A weak version of Cramér's theorem states that under the *weak Cramér's condition:*

$$\forall y \in \mathcal{Y}, \exists \lambda > 0, \int_\Omega e^{\lambda\langle y, \varphi(\omega)\rangle}\, R(d\omega) < \infty \tag{2.2}$$

the empirical means $1/n \sum_{i=1}^n \varphi(Z_i)$ obey the LDP in $\mathcal{X}$ for the topology $\sigma(\mathcal{X}, \mathcal{Y})$ with the good rate function

$$J(x) = \sup(\Delta_x), x \in \mathcal{X} \tag{2.3}$$

where (Δ_x) is the following optimization problem

$$\text{maximize } y \mapsto \langle x, y\rangle - \log \int_\Omega e^{\langle y, \varphi\rangle}\, dR, y \in \mathcal{Y} \tag{2.4}$$

The proof of this result is obtained using Cramér's theorem in $\mathbb{R}^d$ with the law $R\circ(\langle y_1, \varphi(\cdot)\rangle, \ldots, \langle y_d, \varphi(\cdot)\rangle)^{-1}$ ([11], Corollary 6.1.6) together with Dawson-Gärtner's theorem on the projective limits of LDPs ([11], Theorem 4.6.9).

2.4. A dual equality

In the particular case where the constraint function satisfies the following *very strong Cramér's condition:*

$$\forall y \in \mathcal{Y}, \langle y, \varphi(\cdot)\rangle \in B \tag{2.5}$$

the application $P \in \mathcal{P}(\Omega) \mapsto \int_\Omega \varphi\, dP \in \mathcal{X}$, where as a definition: $\langle \int_\Omega \varphi\, dP, y\rangle_{\mathcal{X},\mathcal{Y}} = \int_\Omega \langle y, \varphi\rangle\, dP$, $\forall y \in \mathcal{Y}$, is $\sigma(\mathcal{P}(\Omega), B)$-$\sigma(\mathcal{X}, \mathcal{Y})$-continuous. It follows from the contraction principle ([11], Theorem 4.2.1) that the rate functions of Cramér's and Sanov's LDPs satisfy $J(x) = \inf\{I(P \mid R); P$ such that $\int_\Omega \varphi\, dP = x\}, x \in \mathcal{X}$. In other words, taking (2.3) into account, under the assumption (2.5), the following dual equality holds

$$\inf(\Pi_x) = \sup(\Delta_x) \tag{2.6}$$

where (Π_x) is the following (primal) optimization problem

$$\text{minimize } P \in \mathcal{P}(\Omega) \mapsto I(P \mid R) \text{ subject to } \int_\Omega \varphi\, dP = x \tag{2.7}$$

whose dual problem is precisely (Δ_x) (see (2.4)).

2.5. Csiszár's I-projection

As a consequence, under (2.5), one obtains the following *existence* result: If $x \in \mathcal{X}$ is such that $\sup(\Delta_x) < \infty$, there exists $P \in \mathcal{P}(\Omega)$ such that $\int_\Omega \varphi\, dP = x$ and $I(P \mid R) < \infty$ (hence $P \ll R$). One may ask which are the minimizers of (Π_x). In practice, this corresponds to a *construction* problem (see the huge literature on the entropy maximum, for instance [1] and the references therein). As $I(\cdot \mid R)$ is inf-compact (a good rate function) for the topology $\sigma(\mathcal{P}(\Omega), B)$, the minimizers are attained. On the other hand, $I(\cdot \mid R)$ is strictly convex and $\{P \in \mathcal{P}(\Omega); \int_\Omega \varphi\, dP = x\}$ is a convex set, therefore the minimizer $P_x = \operatorname{argmin}(\Pi_x)$ is unique.

Definition 2.1 (Csiszár, [8]). *P_x is the* I-projection *of R on the convex set $\{P \in \mathcal{P}(\Omega); \int_\Omega \varphi\, dP = x\}$.*

3. Applications to stochastic processes

Clearly, condition (2.5) is very restrictive. Nevertheless, many interesting problems do not violate it. In this section, such examples are presented.

Let us take $\Omega = C([0,1], \mathbb{R}^d)$: the space of continuous paths on $[0,1]$ in $\mathbb{R}^d$. Our reference probability measure R is a process law: $R \in \mathcal{P}(\Omega)$. For any $P \in \mathcal{P}(\Omega)$ and $0 \le t \le 1$, let us denote $P_t \in \mathcal{P}(\mathbb{R}^d)$ the t-marginal of P : the law of the position at time t.

3.1. Schrödinger's bridges

Our aim is to build a process $P \in \mathcal{P}(\Omega)$ such that $P \ll R$, $P_0 = x_0$ and $P_1 = x_1$ where P_0 and P_1 are the initial and final laws of the process and x_0, x_1 are prescribed probability measures on $\mathbb{R}^d$. These constraints are properly described by the constraint function $\varphi : \omega = (\omega_t)_{0 \le t \le 1} \in \Omega \mapsto (\delta_{\omega_0}, \delta_{\omega_1}) \in \mathcal{X}$ where $\mathcal{X} = \mathcal{P}(\mathbb{R}^d) \times \mathcal{P}(\mathbb{R}^d)$ is in separating duality with $\mathcal{Y} = B(\mathbb{R}^d) \times B(\mathbb{R}^d)$ or $\mathcal{Y} = C_o^\infty(\mathbb{R}^d) \times C_o^\infty(\mathbb{R}^d)$. We denote $B(\mathbb{R}^d)$ the space of numerical bounded functions on $\mathbb{R}^d$ and $C_o^\infty(\mathbb{R}^d)$ the space of infinitely differentiable numerical functions on $\mathbb{R}^d$ with a compact support. Indeed, for any $(y_0, y_1) \in B(\mathbb{R}^d) \times B(\mathbb{R}^d)$, $\langle \int_\Omega \varphi\, dP, (y_0, y_1) \rangle_{\mathcal{X},\mathcal{Y}} = \int_\Omega [y_0(\omega_0) +$

$y_1(\omega_1)]\, P(d\omega) = \int_{\mathbb{R}^d} y_0\, dP_0 + \int_{\mathbb{R}^d} y_1\, dP_1$. Hence, $\int_\Omega \varphi\, dP = (P_0, P_1)$. The very strong Cramér's condition (2.5) clearly holds. As the dual equality (2.6) holds, an existence criterion on $x = (x_0, x_1) \in \mathcal{X}$ for such a bridge is: $\sup(\Delta_x) < \infty$. This means

$$\begin{aligned} & \sup_{y_0, y_1 \in B(\mathbb{R}^d)} \left\{ \int_{\mathbb{R}^d} y_0\, dx_0 + \int_{\mathbb{R}^d} y_1\, dx_1 - \log \int_\Omega e^{y_0(\omega_0) + y_1(\omega_1)}\, R(d\omega) \right\} \\ = & \sup_{y_0, y_1 \in C_o^\infty(\mathbb{R}^d)} \left\{ \int_{\mathbb{R}^d} y_0\, dx_0 + \int_{\mathbb{R}^d} y_1\, dx_1 - \log \int_\Omega e^{y_0(\omega_0) + y_1(\omega_1)}\, R(d\omega) \right\} \\ = & \sup_{y_0, y_1 \in C_o^\infty(\mathbb{R}^d)} \left\{ \int_{\mathbb{R}^d} y_0\, dx_0 + \int_{\mathbb{R}^d} y_1\, dx_1 - \log \int_{\mathbb{R}^d \times \mathbb{R}^d} e^{y_0(a) + y_1(b)}\, R_{01}(dadb) \right\} \\ < & \ \infty \end{aligned}$$

where $R_{01}$is the joint law of (ω_0, ω_1) under R. The first equality follows from (2.3) with two different choices of $\mathcal{Y}$: $\mathcal{Y}$ has only to separate $\mathcal{X}$ and satisfy (2.5).
Note that this criterion of existence for a bridge is equivalent to the criterion of existence of a joint law P_{01} with marginal laws x_0 and x_1 such that $I(P_{01} \mid R_{01}) < \infty$.

3.2. Nelson's diffusion processes

Our aim is to build a process $P \in \mathcal{P}(\Omega)$ such that $P \ll R$, $P_t = x_t$ for all $0 \le t \le 1$ and $(x_t)_{0 \le t \le 1} \in \mathcal{X} = C([0,1], \mathcal{P}(\mathbb{R}^d))$ is a prescribed flow of t-marginals. These constraints are properly described by the constraint function $\varphi : \omega = (\omega_t)_{0 \le t \le 1} \in \Omega \mapsto (\delta_{\omega_t})_{0 \le t \le 1} \in \mathcal{X}$. The very strong Cramér's condition (2.5) clearly holds. Let us choose $\mathcal{Y} = C_o^\infty(]0,1[, \mathbb{R}^d)$ with the duality bracket $\langle x, y \rangle = \int_{[0,1] \times \mathbb{R}^d} y(t,a)\, x_t(da) dt$. The existence criterion: $\sup(\Delta_x) < \infty$ writes as follows:

$$\sup_{y \in C_o^\infty(]0,1[, \mathbb{R}^d)} \left\{ \int_{[0,1] \times \mathbb{R}^d} y(t,a)\, x_t(da) dt - \log \int_\Omega R(d\omega) \int_0^1 e^{y(t, \omega_t)}\, dt \right\} < \infty$$

To make things easier, let us take for R the Wiener measure with initial law R_0. In [6], the following results have been proved.

Results 3.1 (Cattiaux & Léonard, [6])

1. *$J(x) < \infty$ implies that there exists $P \ll R$ such that $P_t = x_t$ for all $0 \le t \le 1$.*

2. *We have $J(x) < \infty$ if and only if there exists a vector field $b_x \in \mathcal{H}_x$ such that*

$$\int_{[0,1] \times \mathbb{R}^d} (\partial_t + b_x \cdot \nabla + \frac{\Delta}{2}) f(t,a)\, x_t(da) dt = 0, \forall f \in C_o^\infty(]0,1[\times \mathbb{R}^d)$$

where $\mathcal{H}_x$ is the closure of $\{\nabla g; g \in C_o^\infty(]0,1[\times\mathbb{R}^d)\}$ in the Hilbert space $L_2([0,1]\times\mathbb{R}^d, \overline{x})$ with $\overline{x}(dtda) = x_t(da)dt$.

Moreover, $J(x) = \frac{1}{2}\int_{[0,1]\times\mathbb{R}^d} |b_x(t,a)|^2\, x_t(da)dt$.

3. *If $J(x) < \infty$, the I-projection P_x is the unique absolutely continuous with respect to R solution to the martingale problem associated with the generator $\partial_t + b_x \cdot \nabla + \frac{\Delta}{2}$.*

- Note that the elements of $\mathcal{H}_x$ may be very irregular.
- The proof of 3.1.1 in [6] is different from the above one.

To make precise the type of information carried by $J(x) < \infty$, we give a short proof of the necessary condition of 3.1.2.
Proof. For all $x \in C([0,1], \mathcal{P}(\mathbb{R}^d))$, we have

$$\begin{aligned} J(x) &= \sup_{y\in C_o^\infty(]0,1[,\mathbb{R}^d)} \left\{ \langle x,y\rangle - \log \int_\Omega e^{\langle y,\varphi(\omega)\rangle}\, R(d\omega)\right\} \\ &= \sup_{y\in C_o^\infty(]0,1[,\mathbb{R}^d)} \left\{ \int_{[0,1]\times\mathbb{R}^d} y\, d\overline{x} - \log \int_\Omega \exp\Big(\int_0^1 y(t,\omega_t)\, dt\Big)\, R(d\omega)\right\} \\ &\geq \sup_{f\in C_o^\infty(]0,1[,\mathbb{R}^d)} \left\{ \int_{[0,1]\times\mathbb{R}^d} -(\partial_t + \frac{\Delta}{2}) f\, d\overline{x} - \frac{1}{2}\int_{[0,1]\times\mathbb{R}^d} |\nabla f|^2\, d\overline{x}\right\}. \end{aligned}$$

For the last inequality, only consider y of the special form:

$$-y_f(t,a) = (\partial_t + \frac{\Delta}{2}) f(t,a) + \frac{1}{2}|\nabla f(t,a)|^2, f \in C_o^\infty(]0,1[,\mathbb{R}^d)$$

and use $\int_\Omega \exp\big(\int_0^1 y(t,\omega_t)\,dt\big)\, R(d\omega) = 1$ (exponential martingale). Denoting $\ell_x(f) = \int_{[0,1]\times\mathbb{R}^d} -(\partial_t + \frac{\Delta}{2}) f\, d\overline{x}$, the previous inequality implies that

$$|\ell_x(f)| \leq (J(x) + \frac{1}{2})\|\nabla f\|_{2,\overline{x}}, f \in C_o^\infty(]0,1[,\mathbb{R}^d)$$

where $\|\cdot\|_{2,\overline{x}}$ is the norm of $L_2([0,1]\times\mathbb{R}^d, \overline{x})$. It comes out that, whenever $J(x) < \infty$, $\ell_x(f)$ only depends on ∇f, that is: $\ell_x(f) = \tilde{\ell}_x(\nabla f)$, and $\tilde{\ell}_x$ is a $\|\cdot\|_{2,\overline{x}}$-continuous linear form on $\nabla C_o^\infty(]0,1[,\mathbb{R}^d)$. Finally, by the Riesz representation theorem, there exists a unique $b_x \in \mathcal{H}_x$ such that $\ell_x(f) = \int_{[0,1]\times\mathbb{R}^d} b_x(t,a)\cdot\nabla f(t,a)\, x_t(da)dt$ for all $f \in C_o^\infty(]0,1[,\mathbb{R}^d)$, which is the desired result.

3.3. About the literature

The problem of Schrödinger's bridges has been settled by E. Schrödinger in 1932 [23], and investigated later by several authors, in particular: S. Bernstein [2] and R. Fortet [15]. A. Beurling [3] brought a solution in the spirit of the proof of the present article. For a stimulating presentation of this problem, see H. Föllmer's lectures in Saint-Flour [14]. The above proof already appeared in the author's paper [17].
The first proof of the existence of a Nelson's diffusion process is due to E. Carlen [7]. This problem has then been investigated my many authors who brought several different solutions. The reader may have a look to [5] and [6] for references on the subject and two distinct solutions of this problem.

4. Under the strong Cramér's condition

In the previous sections, the dual equality (2.6) has only been proved under the very strong Cramér's condition (2.5), while Cramér's theorem in $\sigma(\mathcal{X},\mathcal{Y})$ holds under a much weaker condition. In this section, Sanov's theorem is slightly extended. As a consequence, the dual equality is recovered via the contraction principle, under the intermediate condition

$$\forall y \in \mathcal{Y}, \int_\Omega e^{\langle y,\varphi(\omega)\rangle}\, R(d\omega) < \infty \tag{4.1}$$

which is called the *strong Cramér's condition.*
In the next section, a wider extension of Sanov's theorem is stated.

4.1. A slight improvement of Sanov's theorem

Let $\mathcal{M}_\tau$ be the space of measurable functions $\{f : \Omega \to \mathbb{R}; \forall\lambda > 0, \int_\Omega e^{\lambda|f|}\, dR < \infty\}$ and $\mathcal{M}_\tau^*$ be its algebraic dual space. One doesn't identify R-almost surely equal functions in $\mathcal{M}_\tau$. Hence, any empirical measure $1/n\sum_{i=1}^n \delta_{z_i}$ belongs to $\mathcal{M}_\tau^*$ for the duality bracket $\langle 1/n\sum_{i=1}^n \delta_{z_i}, f\rangle = 1/n\sum_{i=1}^n f(z_i), f \in \mathcal{M}_\tau$.

Proposition 4.1 ([19]). *The sequence* $1/n\sum_{i=1}^n \delta_{Z_i}$ *obeys the* LDP *in* $\mathcal{P}(\Omega)\cap\mathcal{M}_\tau^*$, *for the topology* $\sigma(\mathcal{P}(\Omega)\cap\mathcal{M}_\tau^*, \mathcal{M}_\tau)$ *with the good rate function* $I(\cdot \mid R)$.

This result is a corollary of the extended Sanov's theorem stated in next the section (Theorem 5.1), see the Remark . A direct proof is as follows. For any $\vec{f} = (f_1,\dots,f_d) \in \mathcal{M}_\tau^d$, $1/n\sum_{i=1}^n \vec{f}(Z_i)$ obeys the LDP in $\mathbb{R}^d$ with the good rate function $\zeta \in \mathbb{R}^d \mapsto \sup_{\lambda\in\mathbb{R}^d}\{\langle\lambda,\zeta\rangle - \log\int_\Omega e^{\langle\lambda,\vec{f}\rangle}\, dR\}$. Taking the projective limit of these LDPs, one obtains the LDP for L_n in $\mathcal{M}_\tau^*$

for the topology $\sigma(\mathcal{M}_\tau^*, \mathcal{M}_\tau)$ with the good rate function (compare with (2.1))

$$H_{\mathcal{M}}(\ell) = \sup_{f \in \mathcal{M}_\tau} \left\{ \langle \ell, f \rangle - \log \int_\Omega e^f \, dR \right\}, \ell \in \mathcal{M}_\tau^*. \tag{4.2}$$

One concludes by checking that $H_{\mathcal{M}} = I(\cdot \mid R)$ (see [19]).
With the contraction principle, one obtains the following

Corollary 4.2 *The dual equality (2.6) holds under the strong Cramér's condition (4.1).*

We are going to state a characterization of the I-projections P_x when φ satisfies the strong Cramér condition (4.1). This result is a particular case of some results of the author ([16], Theorems 4.4 and 4.5). Let us first introduce the relevant notion of force field.

4.2. Admissible force fields

Let $\mathcal{J}$ be a totally ordered countable index set which admits a smaller element: $\flat$. We consider a family $(n) = (n^j)_{j \in \mathcal{J}}$ of measurable linear forms on $\mathcal{X}$. For any $j \in \mathcal{J}$, let us denote $\mathcal{T}_+^j = \{\langle n^j, \varphi \rangle > 0\} \bigcap \cap_{i<j} \{\langle n^i, \varphi \rangle = 0\}$ and $\mathcal{T}_-^j = \{\langle n^j, \varphi \rangle < 0\} \bigcap \cap_{i<j} \{\langle n^i, \varphi \rangle = 0\}$ with the convention: $\bigcap_{i<\flat} \{\langle n^i, \varphi \rangle = 0\} = \Omega$, so that $\mathcal{T}_+^\flat = \{\langle n^\flat, \varphi \rangle > 0\}$ and $\mathcal{T}_-^\flat = \{\langle n^\flat, \varphi \rangle < 0\}$. We define

$$\mathcal{S} = \bigcap_{j \in \mathcal{J}} \{\langle n^j, \varphi \rangle = 0\}, \quad \mathcal{T}_+ = \bigcup_{j \in \mathcal{J}} \mathcal{T}_+^j \quad \text{and} \quad \mathcal{T}_- = \bigcup_{j \in \mathcal{J}} \mathcal{T}_-^j.$$

Up to a R-negligible set, $\mathcal{S}, \mathcal{T}_+$ and $\mathcal{T}_-$ form a measurable partition of Ω. Let us introduce a notation for the force fields. Let $\overline{z}$ be a measurable linear form on $\mathcal{X}$ and $(n) = (\langle n^j, \varphi \rangle)_{j \in \mathcal{J}}$ as above. We define the application $\langle \overline{z} + \infty \cdot (n), \varphi \rangle : \Omega \to [-\infty, +\infty]$, for any $\omega \in \Omega$, by

$$\langle \overline{z} + \infty \cdot (n), \varphi(\omega) \rangle = \begin{cases} +\infty & \text{if } \omega \in \mathcal{T}_+ \\ -\infty & \text{if } \omega \in \mathcal{T}_- \\ \langle \overline{z}, \varphi(\omega) \rangle & \text{if } \omega \in \mathcal{S}. \end{cases}$$

It is a measurable application. If $(n) = 0$, $\overline{z} + \infty \cdot (n) = \overline{z}$ has no infinite value.

Definition 4.3 ([16]). *One says that* $\overline{z} + \infty \cdot (n)$ *is an* admissible force field *if:*

1. *$\int_{\mathcal{S}}[\langle\overline{z},\varphi\rangle_+ e^{\langle\overline{z},\varphi\rangle_+} + \langle\overline{z},\varphi\rangle_-]\,dR < \infty$ and for all $\varepsilon > 0$, $K \geq 1$, $1 \leq k \leq K$, all functions f_k such that $\int_\Omega |f_k|\log(|f_k|)\,dR < \infty$ and all $g_k \in L_\infty(R)$, there exists $y \in \mathcal{Y}$ such that $\left|\int_\Omega \langle\overline{z}-y,\varphi\rangle[f_k\mathbf{1}_{\langle\overline{z},\varphi\rangle\geq 0} + g_k\mathbf{1}_{\langle\overline{z},\varphi\rangle\leq 0}]\,dR\right| \leq \varepsilon$.*
2. *For all $j \in \mathcal{J}$, $\int_{\cap_{i<j}\{\langle n^i,\varphi\rangle=0\}} \langle n^j,\varphi\rangle_-\,dR < \infty$.*
3. *$R(\mathcal{T}_+) = 0$ and $R(\mathcal{T}_-) < \infty$.*

The subscripts $+$ and $-$ stand for the nonnegative and nonpositive parts of the functions.

4.3. Characterization of the I-projections

An element x of $\mathcal{X}$ is said to be an *admissible constraint*, if $J(x) < \infty$. Being the domain of a convex function, the set of all admissible constraints is a convex subset of $\mathcal{X}$.

Theorem 4.4 (Characterization of the I-projections, [16]). *Let us assume that the strong Cramér's condition (4.1) holds and the σ-field $\mathcal{A}$ on Ω is R-complete.*

1. *For any admissible constraint x_o, there exists an admissible force field $z_{x_o} = \overline{z}_{x_o} + \infty \cdot (n)_{x_o}$ such that*
$$x_o = \int_\Omega \varphi e^{\langle z_{x_o},\varphi\rangle}\,dR \quad \text{and} \tag{4.3}$$
$$J(x_o) = I(e^{\langle z_{x_o},\varphi\rangle}\cdot R \mid R) < \infty. \tag{4.4}$$
Conversely, if x_o is associated with an admissible force field z_{x_o} by formula (4.3), then (4.4) holds.

2. *If x_o is an admissible constraint, then the minimization problem (Π_{x_o}) has a unique solution P_{x_o} in $\mathcal{P}(\Omega)\cap\mathcal{M}_\tau^*$: the set of probability measures which integrate all the functions $\langle y,\varphi(\cdot)\rangle$, $y \in \mathcal{Y}$. The shape of this solution is*
$$P_{x_o} = e^{\langle z_{x_o},\varphi\rangle}\cdot R \tag{4.5}$$
where z_{x_o} is an admissible force field.

Conversely, if $\overline{z} + \infty\cdot(n)$ is an admissible force field, putting $x_o = \int_\Omega \varphi e^{\langle\overline{z}+\infty\cdot(n),\varphi\rangle}\,dR$, we have $J(x_o) < \infty$, $P_{x_o} \triangleq e^{\langle\overline{z}+\infty\cdot(n),\varphi\rangle}\cdot R$ integrates all the functions $\langle y,\varphi(\cdot)\rangle$, $y \in \mathcal{Y}$ and is the unique solution of (Π_{x_o}).

If x_o stands in the relative geometric interior of the effective domain of J, $z_{x_o} = \overline{z}_{x_o}$ has no infinite component. If it stands on the geometric boundary of the effective domain of J, the field of ordered collections of outward normal vectors $(n)_{x_o}$ characterizes the minimal face of the boundary on which x_o stands. Note that x_o is in the relative geometric interior of this face and $\overline{z}_{x_o}$ characterizes x_o in this face.
In situations where $\Omega = C([0,1], \mathbb{R}^d)$, (4.5) is a Girsanov's formula and $e^{\langle \infty \cdot (n)_{x_o}, \varphi \rangle} = \mathbf{1}_{\mathcal{S}}$ (R-almost surely) is the indicator function of a set of paths with finite energy.

4.4. About the literature

A characterization of the minimizer P_x in terms of the cancelation of a gradient is given in [21] and extended in ([20], Theorem 8.10) and ([25], Theorem 2). It doesn't lead to the exact shape of the density of the minimizer: $\frac{dP_x}{dR}$.
A necessary condition for a density to be $\frac{dP_x}{dR}$, and a sufficient condition, are stated in ([8], Theorem 3.1). Except for a finite number of moment constraints, it remains a gap between these conditions to be simultaneously necessary and sufficient. Similar conditions in more general situations are obtained in ([9], Lemma 3.4) and ([20], Theorem 8.20).
For a finite number of qualified constraints, the characterization of $\frac{dP_x}{dR}$ is given in [4] and extended in [10]. Let us mention that a qualified constraint is "interior" and it follows from our results that the force field associated with P_x doesn't take any infinite values.
Theorem 4.4 closes the problem of the characterization of $\frac{dP_x}{dR}$ under the strong Cramér's condition (4.1), without any topological restrictions. It definitely improves the already published related results.

5. Under the weak Cramér's condition

Cramér's theorem holds under the weak Cramér's condition (2.2) while we have only proved Sanov's theorem under the strong Cramér's condition (4.1) (see Proposition 4.1). It is interesting to ask how to extend Sanov's theorem in order to recover Cramér's theorem with the contraction principle, under the weak Cramér's condition. This result has been obtained in collaboration with J. Najim [19], it is stated below. Let us begin illustrating the situation with a simple example .

5.1. Csiszár's example

This example has been studied by I. Csiszár in [9]. Take $\Omega = [0, \infty[$,

$R(d\omega) = c\frac{e^{-\tilde{\lambda}\omega}}{1+\omega^3}\,d\omega$ where $\tilde{\lambda} > 0$, c is the unspecified normalizing constant and $d\omega$ is the Lebesgue measure on $[0,\infty[$. Our constraint function is $\varphi(\omega) = \omega$, $\omega \geq 0$. Note that (2.2) holds but (4.1) fails. The log-Laplace transform of R is $\Lambda(\lambda) = \log \int_{[0,\infty[} e^{\lambda\omega}\,R(d\omega)$. Its effective domain is $]-\infty, \tilde{\lambda}]$ and its left derivative at $\tilde{\lambda} : \Lambda'(\tilde{\lambda}) \triangleq \tilde{x}$ is finite. Therefore, its convex conjugate $J = \Lambda^*$ has effective domain $]0,\infty[$ and is affine with slope $\tilde{\lambda}$ on $[\tilde{x},\infty[$. The problem (Π_x) consists of minimizing $P \mapsto I(P \mid R)$ under the constraint $\int_{[0,\infty[} \omega\,P(d\omega) = x$. In [9], it is shown that

- for any $x > 0$, the dual equality: $\inf\{I(P \mid R); \int_{[0,\infty[} \omega\,P(d\omega) = x\} = J(x)$, holds

- for any $0 < x \leq \tilde{x}$, the infimum is attained at $P_x(d\omega) = ce^{\lambda_x\omega}\,R(d\omega)$ where $\Lambda'(\lambda_x) = x$

- for any $x > \tilde{x}$, the infimum is not attained, but any minimizing sequence (P_n) (i.e. $\lim_{n\to\infty} I(P_n \mid R) = J(x)$ and $\int_{[0,\infty[} \omega\,P_n(d\omega) = x, \forall n \geq 1$), converges in the sense of total variation to $\tilde{P}(d\omega) \triangleq P_{\tilde{x}}(d\omega) = ce^{\tilde{\lambda}\omega}\,R(d\omega) = c\frac{d\omega}{1+\omega^3}$.

I. Csiszár introduced in [9] the notion of *generalized I-projection* to take this phenomenon into account.

5.2. The extended Sanov's theorem

Let $\mathcal{L}_\tau$ stand for the space of measurable functions on Ω which admit "some finite exponential moment":

$$\mathcal{L}_\tau = \{f : \Omega \to \mathbb{R}; \exists\lambda > 0, \int_\Omega e^{\lambda|f|}\,dR < \infty\}$$

Its algebraic dual is denoted $\mathcal{L}_\tau^*$. Proceeding as in the proof of Proposition 4.1, one obtains by projective limits of Cramér's theorem in $\mathbb{R}^d$ that $\{L_n; n \geq 1\}$ obeys the LDP in $\mathcal{L}_\tau^*$ for the topology $\sigma(\mathcal{L}_\tau^*, \mathcal{L}_\tau)$ and with the good rate function

$$H_{\mathcal{L}}(\ell) = \sup_{f\in\mathcal{L}_\tau}\left\{\langle \ell, f\rangle - \log\int_\Omega e^f\,dR\right\}, \ell \in \mathcal{L}_\tau^*$$

(compare with (2.1) and(4.2)). As $\mathcal{M}_\tau \subset \mathcal{L}_\tau$, we have $H_{\mathcal{L}}(\ell) \geq H_{\mathcal{M}}(\ell')$ $= \begin{cases} I(\ell' \mid R) & \text{if } \ell' \in \mathcal{P}(\Omega) \\ \infty & \text{otherwise} \end{cases}$ where ℓ' is the restriction of ℓ to $\mathcal{M}_\tau$.

In [19], it is proved that $H_{\mathcal{L}}(\ell) < \infty$ implies that ℓ matches with an element of the topological dual space L'_τ of the Orlicz space $L_\tau \triangleq \mathcal{L}_\tau/R$-a.s. : the factor space of $\mathcal{L}_\tau$ for the R-a.s. equality, endowed with the Luxemburg norm $\|f\|_\tau = \inf\{a > 0; \int_\Omega \tau(f/a)\, dR \le 1\}$ where τ is the Young function $\tau(s) = e^{|s|} - |s| - 1$, $s \in \mathbb{R}$.
Any element $\ell \in L'_\tau$ is decomposed into $\ell = \ell^a + \ell^s$, the sum of its absolutely continuous part $\ell^a = \frac{d\ell^a}{dR} \cdot R$ with $\frac{d\ell^a}{dR} \in L_{\tau^*}$ and of its singular part $\ell^s \in L^s_\tau$. The space L_{τ^*} is the Orlicz space associated with the convex conjugate τ^* of $\tau : \tau^*(t) = (|t|+1)\log(|t|+1) - |t|$, $L_{\tau^*} = \{g : \Omega \to \mathbb{R}; \int_\Omega \tau^*(g)\, dR < \infty\}$. The space L^s_τ of singular forms consists of all $\ell \in L'_\tau$ such that $\langle \ell, f\rangle = 0$, for all $f \in \mathcal{M}_\tau$. Therefore, one can write $L'_\tau \simeq (L_{\tau^*} \cdot R) \oplus L^s_\tau$.
Let $\mathcal{Q}$ be the set of all $\ell \in \mathcal{L}^*_\tau$ which are nonnegative: $\langle \ell, f\rangle \ge 0, \forall f \ge 0$, with unit mass: $\langle \ell, \mathbf{1}\rangle = 1$. Note that since $\mathbf{1}$ belongs to $\mathcal{M}_\tau$, we have $\langle \ell, \mathbf{1}\rangle = 0$ for all $\ell \in L^s_\tau$. It comes out that except for 0, the elements of L^s_τ cannot be represented as measures (they are finitely additive, but not σ-additive set functionals).

Theorem 5.1 (Extended Sanov's theorem, [19]). *The sequence of random empirical measures $\{L_n; n \ge 1\}$ obeys the* LDP *in $\mathcal{Q}$ for the topology $\sigma(\mathcal{Q}, \mathcal{L}_\tau)$ with the good rate function*

$$I(\ell) = \begin{cases} I_a(\ell^a) + I_s(\ell^s) & \text{if } \ell \in L'_\tau \\ \infty & \text{otherwise} \end{cases}$$

where $\ell = \ell^a + \ell^s$, $I_a(\ell^a) = \begin{cases} I(\ell^a \mid R) & \text{if } \ell^a \in \mathcal{P}(\Omega) \\ \infty & \text{otherwise} \end{cases}$ *and* $I_s(\ell^s) = \sup\{\langle \ell^s, f\rangle; f \in \mathcal{L}_\tau, \int_\Omega e^f\, dR < \infty\}$.

Proof. See [19]. Note that if $I_s(\ell^s) < \infty$, then ℓ^s is nonnegative. The function I_s is the recession function of I_a, it is also the support functional of the convex set $\{f \in \mathcal{L}_\tau; \int_\Omega e^f\, dR < \infty\}$.
Remark . One recovers Proposition 4.1 contracting the LDP of Theorem 5.1 with the application which associates with any element of $\mathcal{L}^*_\tau$ its restriction to $\mathcal{M}_\tau$. Indeed, the restriction to $\mathcal{M}_\tau$ of any singular form (in $\mathcal{L}^s_\tau$) is zero.

The equality $\langle \ell, \varphi\rangle = x$ with $\ell \in \mathcal{Q}$ and $x \in \mathcal{X}$ means $\langle \ell, \langle y, \varphi\rangle\rangle = \langle x, y\rangle$, for all $y \in \mathcal{Y}$. In view of the continuity of $\ell \in \mathcal{Q} \mapsto \langle \ell, \varphi\rangle \in \mathcal{X}$ with respect to the topologies $\sigma(\mathcal{Q}, \mathcal{L}_\tau)$ and $\sigma(\mathcal{X}, \mathcal{Y})$, the contraction principle yields the following

Corollary 5.2 *Let us consider the following extension $(\overline{\Pi}_x)$ of (Π_x) :*

$$\text{minimize } \ell \in \mathcal{Q} \cap L'_\tau \mapsto I(\ell^a \mid R) + I_s(\ell^s) \text{ subject to } \langle \ell, \varphi\rangle = x \qquad (5.1)$$

The dual equality $\sup(\Delta_x) = \inf(\overline{\Pi}_x)$ *holds under the weak Cramér's condition (2.2).*

Since I is a good rate function, when $J(x)(:= \sup(\Delta_x)) < \infty$, the infimum is attained for $(\overline{\Pi}_x)$, although it may not be attained for (Π_x).
As I_s is positively homogeneous, it is not strictly convex and $(\overline{\Pi}_x)$ may admit several solutions.

Definition 5.3 *The minimizers of* $(\overline{\Pi}_x)$ *are called* entropic projections.

5.3. Back to Csiszár's example

Now, we consider $(\overline{\Pi}_x)$ instead of (Π_x) in the above Csiszár's example. One can prove ([18]) that ℓ_x is a solution to $(\overline{\Pi}_x)$ if and only if $\ell_x = \begin{cases} P_x & \text{si } x \le \tilde{x} \\ \tilde{P} + (x-\tilde{x})\xi & \text{si } x \ge \tilde{x} \end{cases}$ where ξ is any nonnegative element of L_τ^s such that

1. the constraint $\langle \xi, \varphi \rangle = 1$ is satisfied and
2. its "support" is determined by $\langle \xi, f \rangle = 0$ for all nonnegative f in L_τ such that there exists $t > 0$ with $\int_{[0,\infty[} e^{\tilde{\lambda}\varphi + tf}\, dR < \infty$.

Note that

- for all $x \le \tilde{x}$, $\langle \ell_x, \varphi \rangle = \langle P_x, \varphi \rangle = x$ and $I(\ell_x) = I(P_x \mid R) = J(x)$ and
- for all $x \ge \tilde{x}$, $I(\ell_x) = I(\tilde{P} \mid R) + \langle (x-\tilde{x})\xi, \tilde{\lambda}\varphi \rangle = J(\tilde{x}) + \tilde{\lambda}(x-\tilde{x}) = J(x)$.

For the second item, one can prove that under the constraint $\langle \xi, \varphi \rangle = 1$, the supremum of $f \mapsto \langle \xi, f \rangle$ subject to $\int_{[0,\infty[} e^f\, dR < \infty$ is attained at $\tilde{\lambda}\varphi$.

5.4. About the literature

Improvements of the usual Sanov's theorem for the topology $\sigma(\mathcal{P}(\Omega), B))$ have been obtained by P. Eischelbacher and U. Schmock ([12], [13]). They are close to Proposition 4.1. A. Schied [22] shows that the relative entropy $I(\cdot \mid R)$ may not be inf-compact for topologies $\sigma(\mathcal{P}(\Omega), \mathcal{F})$ when $\mathcal{F}$ is not included in $\mathcal{M}_\tau$.
The dual equality in Corollary 5.2 is proved in [17] with convex analysis. In [18], the author characterizes the minimizers of (5.1). This extends Theorem 4.4.

In [19], Theorem 5.1 is exploited to improve the Gibbs conditioning principle obtained by D. Stroock and O. Zeitouni [24] (see also: [11] for a detailed presentation, and [9] for an alternate statement).

References

[1] 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.

[2] S. Bernstein. Sur les liaisons entre les grandeurs aléatoires *Vehr. des intern. Mathematikerkongr., Zürich 1932, Band 1.*

[3] A. Beurling. An automorphism of product measures *Ann. Math.*, 72:189–200, 1960.

[4] J. M. Borwein and A. S. Lewis. Partially-finite programming in L_1 and the existence of maximum entropy estimates *SIAM J. Optim.*, 3:248–267, 1993.

[5] P. Cattiaux and C. Léonard. Minimization of the Kullback information of diffusion processes. *Ann. Inst. Henri Poincaré*, 30:83–132, 1994.

[6] P. Cattiaux and C. Léonard. Large deviations and Nelson processes. *Forum Math.*, 7:95–115, 1995.

[7] E. Carlen. Conservative diffusions. *Comm. Math. Phys.*, 94:293–315, 1984.

[8] I. Csiszár. I-Divergence Geometry of Probability Distributions and Minimization Problems. *Ann. Probab.*, 3:146–158, 1975.

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

[10] I. Csiszár, F. Gamboa and E. Gassiat. MEM pixel correlated solutions for generalized moment and interpolation problems, *IEEE Trans. Inform. Theory.*, 45:2253–2270, November 1999.

[11] A. Dembo and O. Zeitouni. Large Deviations Techniques and Applications, Second edition. *Springer-Verlag*, 1998.

[12] P. Eischelsbacher and U. Schmock. Large and moderate deviations of products of empirical measures aud U-empirical processes in strong topologies. *preprint*, 1997.

[13] P. Eischelsbacher and U. Schmock. Exponential approximations in completely regular topological spaces and extensions of Sanov's theorem. *Stochastic Processes and their Applications*, 77:233–251, 1998.

[14] H. Föllmer. Random Fields and Diffusion Processes. Cours à l'Ecole d'Été de Probabilités de Saint-Flour. *Lecture Notes in Math. 1362, Springer Verlag*, 1988.

[15] R. Fortet. Résolution d'un système d'équations de M. Schrödinger. *J. Math. Pures. Appl.*, IX:83–105, 1940.

[16] C. Léonard. Minimizers of energy functionals, *Preprint Université d'Orsay*, 1996.

[17] C. Léonard. Minimization of energy functionals applied to some inverse problems, *Preprint Université d'Orsay*, 1996.

[18] C. Léonard. Minimizers of energy functionals under not very integrable constraints *Preprint Ecole Polytechnique, CMAP*, 2000.

[19] C. Léonard and J. Najim. An extension of Sanov's theorem. Application to the Gibbs conditioning principle. *pPreprint Ecole Polytechnique, CMAP*, 2000.

[20] F. Liese and I. Vajda. Convex Statistical Distances. *Leipzig: B.G. Teubner*, 1987.

[21] L. Rüschendorf. On the minimum discrimination information theorem. *Statist. Decisions*, Supplementary volume 1, 1984.

[22] A. Schied. Cramér's condition and Sanov's theorem. *Statitics and Probability Letters*, 39:55–60, 1998.

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

[24] D.W. Stroock and O. Zeitouni. Microcanonical distributions, Gibbs states, and the equivalence of ensembles. In R. Durrett and H. Kesten, editors, Festchrift in honour of F. Spitzer. *Birkhäuser. Basel, Switzerland*, 399–424, 1991.

[25] M. Teboulle and I. Vajda. Convergence of Best ϕ-Entropy Estimates. *IEEE Trans. Inform. Theory*, 39:297–301, January 1993.

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

Centre de Mathématiques Appliquées, Ecole Polytechnique
91128 Palaiseau Cedex, France

leonard@modalx.u-paris10.fr