The Annals of Statistics
1997, Vol. 25, No. 1, 328–350

BAYESIAN METHODS AND MAXIMUM ENTROPY FOR ILL-POSED INVERSE PROBLEMS

BY F. GAMBOA[1] AND E. GASSIAT[2]

Université Paris Sud and Université d'Evry-Val d'Essonne

In this paper, we study linear inverse problems where some generalized moments of an unknown positive measure are observed. We introduce a new construction, called the maximum entropy on the mean method (MEM), which relies on a suitable sequence of finite-dimensional discretized inverse problems. Its advantage is threefold: It allows us to interpret all usual deterministic methods as Bayesian methods; it gives a very convenient way of taking into account prior information; it also leads to new criteria for the existence question concerning the linear inverse problem which will be a starting point for the investigation of superresolution phenomena. The key tool in this work is the large deviations property of some discrete random measure connected with the reconstruction procedure.

1. Introduction.

1.1. *The inverse problem.* In this paper, we study the inverse problem

$$Y = \Phi \cdot \mu + \varepsilon, \tag{1.1}$$

where $Y = (Y_1, \dots, Y_k)$ is the finite-dimensional observation, μ is an unknown infinite-dimensional parameter, Φ is a known linear operator (highly noninvertible; k is fixed) and $\varepsilon = (\varepsilon_1, \dots, \varepsilon_k)$ is a multidimensional noise. The inverse problem consists in recovering μ on the basis of the observations Y. Many physical problems may be formalized in this way, for instance, in tomography, spectroscopy or astronomy [see, e.g., McLaughlin (1984)]. In most applications, some prior information is available. We focus here on the generic (representative) moment problem, where μ is a positive measure on a space U and Φ is a k-dimensional moment operator

$$\Phi \cdot \mu = \left(\int_U \phi_j \, d\mu \right)_{j=1,\dots,k}. \tag{1.2}$$

We propose a new construction, issuing from maximum entropy ideas, which we call the maximum entropy on the mean method (MEM). First of all, MEM is not exactly what should be called a new statistical method, but a new

Received February 1993; revised December 1995.
[1]Also affiliated with Université Paris-Nord, Institut Galilée, 93430 Villetaneuse, France.
[2]Also affiliated with Laboratoire de Statistiques, Université Paris Sud, 91405 Orsay, France.
AMS 1991 *subject classifications*. 62A99, 60A99.
Key words and phrases. Moment problems, large deviations, Bayesian statistics.

stochastic procedure which leads, when some discretization step tends to 0, to deterministic methods. We have in mind three principal goals:

1. In a deterministic framework or in a pure statistical one, regularization methods are very popular. The choice of the regularization criterion is left to the user. It is then interesting to give a Bayesian interpretation of these criteria, in order to suggest a rational choice between them. Let us recall the results concerning the regression problem $Y_i = f(x_i) + \varepsilon_i$, $i = 1, \dots, k$. (This is a particular moment problem, the interpolation operator being the integral of (1) the product of the indicator function of an interval and (2) the derivative of f.) The regularization methods minimize a criterion $\|Y - \Phi f\|^2 + \lambda J(f)$, where λ is a parameter and J is a suitable functional; see Wahba (1990) or Barry (1986), for instance. A Bayesian interpretation can be given using the distribution of Gaussian processes as priors on the function f, as Wahba shows, for instance. All these Gaussian priors lead to criteria which are functionals of derivatives of f (Sobolev norms): this is a restrictive way to take into account many kinds of prior information and even for regularity it seems restrictive. Our construction allows us to clarify the correspondence between Bayesian rules and regularization technique, including of course the most popular ones, such as least squares, Shannon entropy, Burg entropy, and L^p-norms. This goal is in spirit similar to that pursued by the paper of Csiszar (1991), where deterministic criteria are related to the different axioms which could be required for the estimation method.
2. The MEM construction gives a natural and practical way of introducing nonlinear but convex constraints considered as prior information. For instance, if the unknown measure (μ) has a density function (with respect to some known prior P) f, the constraint may be a qualitative shape $a(x) \le f(x) \le b(x)$, an energy constraint $\int f^2(x)\,dP(x) \le 1$ or a positivity constraint $f \ge 0$. In this paper, we focus on the positivity constraint, whereas shape constraints or energy constraints appear in Dacunha-Castelle and Gamboa (1990) and in Gamboa and Gassiat (1991).
3. Our only concern are truly ill-posed problems, so that the classical asymptotics governed by the consistency of an estimator when the number of observations increases is outside of the scope of this paper; such consistency questions concern problems completely different in spirit. We are, on the other hand, interested in the following asymptotic analysis linked with what is called superresolution by people working in physical applications, such as optics or astronomy. Considering nonnoisy observations $Y = \Phi \cdot \mu$, a boundary set (which will be made precise later on) separates data Y for which the inverse problem has solutions (in general, infinitely many solutions) and data Y for which the inverse problem has no solution when positivity is taken into account. If we are near this boundary (or if the noise moves the inverse problem a little away from the boundary), all solutions built using nonlinear methods (that incorporate the positivity) are very close. The use of nonlinear "positive" methods drastically im-

proves the solution, compared to linear methods. We are then concerned with the following asymptotic: what happens when the data are close to the boundary.

Let us now give the precise setting of the MEM construction and how it allows answers concerning the three previous aims.

1.2. *The maximum entropy on the mean construction and sequences of Bayesian problems*. Maximum entropy on the mean (MEM) is a construction where the inverse problem (1.1) is approximated by a sequence of finite-dimensional problems, which are obtained by a discretization of the underlying space U. The MEM estimator is then obtained as the limit of the discretized estimators defined for the finite-dimensional problems. Notice that the number of observations stays finite, equal to k, and the asymptotic only concerns the space discretization.

The first step is to choose a given probability P on U which can be thought of as a translation of prior information on μ. Let $(x_i)_{i \in \mathbb{N}}$ be a given deterministic sequence of U such that the discrete measure $P_n := (1/n)\sum_{i=1}^n \delta_{x_i}$ converges weakly to this chosen probability measure P.

The second step reflects another kind of prior information. To each coordinate x_i is associated a random variable Z_i. Let F_n be the distribution of $Z^n = (Z_1, \dots, Z_n)$. Suppose for instance that the prior information is $a(x) \leq f(x) \leq b(x)$, where $f(x)\,dx$ is the probability to be reconstructed. Then, for every i, we choose the support of the distribution F_n^i of Z_i to be included in $[a(x_i), b(x_i)]$ and if we have no more information, $F_n = \otimes_{i=1}^n F_n^i$. If the prior information is $\int f^2(x)\,dx \leq 1$, we choose F_n with a support included in the unit ball of $\mathbb{R}^n$.

Let us now introduce two kinds of estimators of the unknown μ when the noise-corrupted moments $Y = \Phi\mu + \varepsilon$ are observed. The observation Y is definitely of fixed dimension k. Let

$$\nu_n := \frac{1}{n} \sum_{i=1}^n Z_i \delta_{x_i}. \tag{1.3}$$

We define the discrete estimators of μ as

$$\hat{\nu}_n^{\text{bay}} = E_{F_n}\left[\nu_n \middle| \left\| \int \phi \, d\nu_n - Y \right\| \leq \rho \right]$$

if ρ quantifies the noise level or, more generally,

$$\hat{\nu}_n^{\text{bay},\mathscr{C}} = E_{F_n}\left[\nu_n \middle| \int \phi \, d\nu_n \in \mathscr{C} \right], \tag{1.4}$$

where $\mathscr{C}$ is a convex compact set in $\mathbb{R}^k$, which describes together the observation and the noise level (for instance, a confidence set). For every level n of the discretization of the space U, $\hat{\nu}_n^{\text{bay},\mathscr{C}}$ is a Bayesian estimator with a priori F_n; we then have a sequence of n-dimensional Bayesian problems with a k-dimensional observation. When n tends to infinity, we shall prove under weak assumptions that $\hat{\nu}_n^{\mathrm{bay},\mathscr{C}}$ is convergent. To solve the linear inverse problem, a direct infinite-dimensional Bayesian reconstruction method like those used usually in statistics [see Ferguson (1974)] seems difficult: to do this, we have to endow $\mathscr{M}^+(U)$, the space of positive measures on U, with a prior Q. Then we have to define the Bayesian estimator by conditioning with the observation $\int_U \phi \, d\mu \in \mathscr{C}$:

$$E_Q\left[\mu \middle| \int_U \phi \, d\mu \in \mathscr{C}\right].$$

The drastic effect of Q is that it introduces an arbitrary prior by lack of invariance. For instance, if no prior is given, one has reasonably to require that Q preserves invariance for the transformations which preserve P. As an example, when P is the Lebesgue measure on $[0,1]$, Q has to be invariant for the rearrangement group (analogous to the permutation group on $\mathbb{N}$). In general there are no Q with this property. So we propose to study the sequence $\hat{\nu}_n^{\mathrm{bay},\mathscr{C}}$ of finite-dimensional Bayesian estimators.

We then define the MEM estimate: at stage n, we choose the distribution P_n^{MEM} of Z^n using a maximum entropy principle

$$K\big(P_n^{\mathrm{MEM}}, F^{\otimes n}\big) = \min_{\|E_R \int \phi \, d\nu_n - Y\| \le \rho} K(R, F^{\otimes n}), \tag{1.5}$$

where $K(\cdot,\cdot)$ is the Kullback information. [For definition and properties of the Kullback information, see, for instance, Dacunha-Castelle and Duflo (1986).] Set

$$\hat{\nu}_n^{\mathrm{MEM}} := E_{P_n^{\mathrm{MEM}}}(\nu_n). \tag{1.6}$$

So for fixed n, $\hat{\nu}_n^{\mathrm{MEM}}$ is the maximum entropy reconstruction of μ with reference measure F_n. Notice that the maximum entropy principle is not applied directly as in previous maximum entropy methods: Roughly speaking, MEM seems to construct a stochastic process of maximum entropy subject to the constraint that its mean function is a solution of inverse problem (1.1) with noise level ρ. However, the coherent system of finite distributions involved in the discretization does not lead to any infinite distribution on the space of positive measures on U by lack of compactness. No direct infinite-dimensional randomization would lead to the same reconstructions. Note also that this point of view is of course completely different from that of papers such as Diaconis and Freedman (1986) which are concerned with classical consistency.

In the following sections, our results show that, under very weak assumptions, when the number n of discretization points tends to infinity, the following statements hold:

1. Reconstruction $\hat{\nu}_n^{\mathrm{MEM}}$ converges to a solution $\hat{\nu}^{\mathrm{MEM}}$ of the inverse problem with noise level ρ (see Theorems 2.1 and 2.2).
2. $\hat{\nu}_n^{\mathrm{bay}}$ satisfies similar asymptotic results and converges to $\hat{\nu}^{\mathrm{bay}}$ (see Theorem 2.3).
3. The Bayesian estimate and the MEM estimate are the same: $\hat{\nu}^{\mathrm{bay}} = \hat{\nu}^{\mathrm{MEM}}$ (see Theorem 2.3).

Moreover, we also prove that the inverse problem has solutions if and only if the asymptotic estimate $\hat{\nu}^{\mathrm{bay}}$ (or $\hat{\nu}^{\mathrm{MEM}}$) exists, and this holds for any choice of distribution F. This leads us to criteria based only on the observations for deciding on the existence of solutions to the inverse problem. In fact, from the Bayesian point of view, regularization functions and existence criteria are linked in a very narrow manner as we shall discuss below. Suppose that as Bayesian n-dimensional a priori measure we use $F^{\otimes n}$ for every n. Then we can associate to F a functional J_F such that $\hat{\nu}^{\mathrm{bay}}$ is exactly obtained as a minimizer of $J_F(\mu)$ subject to the constraint $\|\int \Phi\, d\mu - Y\| \leq \rho$. Thus regularization methods have a very clear Bayesian interpretation if one considers that Bayesian means increasing finite-dimensional Bayesian problems. Let us give some examples. Let $\mu = f \cdot P + \sigma$ be the decomposition of μ as the sum of the absolutely continuous part with respect to P and the singular one. In the most simple cases, J_F is the nonlinear integral functional associated with the Cramér transform of the distribution F [see (2.7)]. Other examples are given in the following list.

1. *Shannon entropy*. $J_F(\mu) = \int f \log f dP - \int f dP + 1$ when $\sigma = 0$, $J_F(\mu) = +\infty$ if $\sigma \neq 0$, corresponds to the case where F is the Poisson distribution with mean 1.
2. *Fermi–Dirac entropy*. $J_F(\mu) = \int f \log f dP + \int (1-f)\log(1-f)\, dP - \log 2$ when $\sigma = 0$, $J_F(\mu) = +\infty$ if $\sigma \neq 0$, corresponds to the case where F is the two point distribution $\frac{1}{2}(\delta_0 + \delta_1)$ and the shape information is $0 \leq f \leq 1$.
3. L_p*-norms*. When F has a density proportional to $\exp(-x^p)$ on $\mathbb{R}_+$, $J_F(\cdot)$ is equivalent to an L^p-norm.
4. *Energy*. $J_F(\mu) = \frac{1}{2}\int f^2\, dP$ when $\sigma = 0$, $J_F(\mu) = +\infty$ if $\sigma \neq 0$, corresponds to the case where F is the standard Gaussian distribution with no information on the meausre. The associated regularization method is least squares.
5. *Burg entropy*. $J_F(\mu) = -\int \log f dP + \alpha\mu(U) - 1$ corresponds to the case where F is an exponential distribution with mean α.
6. α*-Burg entropy*. If F is the Poissonized convolution of the $\gamma(\beta, \alpha)$ distribution, then

$$J_F(\mu) = -\beta^{(\beta+1)^{-1}}\left(\alpha + \frac{\alpha^{\beta}}{\beta}\right)\int f^{1-(\beta+1)^{-1}}\, dP + \alpha\sigma(U).$$

So the most popular regularization functionals are covered by our results. Some of the previous examples are introduced in detail in Dacunha-Castelle and Gamboa (1990), and item 4 above is related to infinite-dimensional a priori measures considered as stochastic fields on U.

In this paper, we focus our attention on the special case where the finite-dimensional prior has form $F^{\otimes n}$ with a technical choice of F, in order to obtain existence criteria. Such criteria are obtained as an immediate consequence of the MEM construction.

These criteria enlighten the problem of singular solutions, but, in fact, the most interesting feature in such a choice of F (as we see later) is that it gives an illuminating explanation of the phenomenon called superresolution in statistical signal theory; see Section 2.4. Let us note that superresolution problems have already interested statisticians such as Frieden (1985), Donoho, Johnstone, Hoch and Stern (1992) and Donoho (1993) [see also Donoho and Gassiat (1992) and Gassiat (1991)]. Here we state it as a boundary inverse problem, so that criteria based on the proximity of the data to the boundary are needed to identify when this phenomenon occurs.

The main tool used in this paper is large deviation theory [Varadhan (1984)]. In the usual parametric models, to prove consistency of the Bayesian estimator, one definitely must use large deviation theory in its weaker form: Laplace methods, where the asymptotic is on the number of observations. Here, to study the convergence of the Bayes estimator $\hat{\nu}_n^{\mathrm{bay}}$ (the asymptotic being on the discretization number), we also use, of course, large deviation theory in a more complicated setting [see also Csiszar (1984) for similar use of large deviations]. If one is not interested in Bayesian interpretations, the whole problem may be investigated using purely deterministic convex methods, which was done by Borwein and Lewis (1993) after a previous version of this work. However, our probabilistic framework allows us to consider a deeper statistical analysis, especially to understand how the distribution of the noise interferes with the Bayesian point of view; see Gamboa (1994).

This paper is organized as follows. In Section 2, we state the assumptions precisely and give the theorems concerning the asymptotic convergence of the discretized reconstruction. In Sections 2.2 and 2.3, we prove the equivalence between Bayesian methods, maximum entropy on the mean methods and deterministic methods. In Section 2.4, we give the existence criteria that are consequences of the large deviations point of view. A subsequent section details different a posteriori behaviors of the discretized estimator. All proofs are collected in Section 3.

2. Convergence of the discretized MEM and Bayesian distributions.

2.1. *Notation and assumptions.* Let U be a given compact set of $\mathbb{R}^q$, $q \geq 1$, or more generally of a Polish space. The set U is endowed with its Borel σ-field $\mathscr{B}(U)$, and P is a given probability measure on U. We will assume that the support of P is exactly U. The terms $\mathscr{M}(U)$, $\mathscr{M}_+(U)$ and $\mathscr{P}(U)$ are, respectively, the spaces of measures, positive measures and probability measures on U, endowed with the weak topology and $C(U)$ is the set of continuous functions on U endowed with the uniform norm. The function $\phi := (1, \phi_2, \dots, \phi_k)$ is a given k-dimensional real-valued P-a.s. continuous function defined on U. The restrictor $\phi_1 = 1$ could be replaced by a more general compactness assumption, but we keep it for the sake of intuition, since it says that the searched measure is nearly a probability measure.

For any convex compact set $\mathscr{C}$ of $\mathbb{R}^k$ define

$$\mathscr{S}(\mathscr{C}) := \left\{ \mu \in \mathscr{M}^+(U) \colon \int_U \phi(x)\, d\mu(x) \in \mathscr{C} \right\}. \tag{2.1}$$

When $\mathscr{C}$ is the closed ball centered on Y with radius ρ, $\mathscr{S}(\mathscr{C})$ is the set of solutions of the inverse problem with noise level ρ. F is a given probability measure on $\mathbb{R}_+$. In this paper we will always assume that the convex hull of the support of F is $\mathbb{R}_+$.

Let us introduce some more notation and assumptions. The function ψ is the log-Laplace transform of F, $D(\psi)$ is its domain and γ is its Legendre transform:

$$\psi(\tau) := \log \int_0^{+\infty} \exp(\tau y)\, dF(y), \qquad \tau \in \mathbb{R}, \tag{2.2}$$

$$D(\psi) := \{\tau \in \mathbb{R}, \psi(\tau) < \infty\}, \tag{2.3}$$

$$\gamma(t) := \sup_{\tau \in \mathbb{R}} (\tau t - \psi(\tau)), \tag{2.4}$$

$$D(F, \phi) := \left\{ v \in \mathbb{R}^k : \int_U \psi(\langle v, \phi(x) \rangle)\, dP(x) < \infty \right\}, \tag{2.5}$$

$$D'(F, \phi) := \{ v \in \mathbb{R}^k : \forall x \in U, \langle v, \phi(x) \rangle \in D(\psi)\}. \tag{2.6}$$

Here, $\langle \cdot, \cdot \rangle$ denotes the usual scalar product on $\mathbb{R}^k$.

Let g be a measurable function on U. We define the nonnegative (possibly infinite) functional

$$\Gamma(g) := \int_U \gamma[g(x)]\, dP(x).$$

For μ in $\mathscr{M}(U)$ with Radon–Nikodym decomposition $\mu = gP + \sigma$ and $\alpha \in \overline{\mathbb{R}}_+$, set

$$J_F(\mu) := \begin{cases} \alpha\sigma(U) + \Gamma(g), & \text{if } \mu \in \mathscr{M}_+(U), \\ +\infty, & \text{otherwise} \end{cases} \tag{2.7}$$

Assumption (H1). The domain $D(\psi)$ is open: $D(\psi) = (-\infty, \alpha)$, $\alpha \in \overline{\mathbb{R}}_+$.

The case $\alpha < \infty$ says that F weights the tail. In this case, our Assumption (H1) on $D(\psi)$ could be weakened with a little extra work to

$$\overline{D(\psi)} = (-\infty, \alpha] \quad \text{and} \quad \lim_{\tau \to \alpha} \psi'(\tau) = +\infty$$

(that is, ψ is essentially smooth [Rockafellar (1970), page 251]). The case where $\alpha = +\infty$ has been studied in Dacunha-Castelle and Gamboa (1990).

Assumption (H2). For all $v \in \mathbb{R}^k \setminus \{0\}$, $P(\Omega_0(v)) < 1$, where

$$\Omega_\eta(v) := \{x \in U : \langle v, \phi(x) \rangle = \eta\}, \qquad v \in \mathbb{R}^k, \eta \in \mathbb{R}.$$

It will be clear later how the use of Laplace methods leads naturally to such an assumption.

Assumption (H3). There exist $g \in L^1(P)$, $g > 0$ P-a.s., $\int_U \phi(x) g(x)\, dP(x) \in \mathscr{C}$.

Assumption (H3) says that the inverse problem has at least one solution. Criteria to decide validity of Assumption (H3) are the object of Section 2.4.

2.2. *The convergence theorem for* $\hat{\nu}_n^{\mathrm{MEM}}$. Before stating our result, let us explain the construction of the discretized estimate $\hat{\nu}_n^{\mathrm{MEM},\mathscr{C}}$ generalizing (1.5) and (1.6) to the case where $\mathscr{C}$ is not necessarily a ball. Define $P_n^{\mathrm{MEM},\mathscr{C}}$ by

$$K\left(P_n^{\mathrm{MEM},\mathscr{C}}, F^{\otimes n}\right) = \min_{E_R \int \phi \, d\nu_n \in \mathscr{C}} K\left(R, F^{\otimes n}\right). \tag{2.8}$$

Set

$$\hat{\nu}_n^{\mathrm{MEM},\mathscr{C}} := E_{P_n^{\mathrm{MEM},\mathscr{C}}}(\nu_n). \tag{2.9}$$

Let us recall the results of Gamboa and Gassiat (1991) concerning $\hat{\nu}_n^{\mathrm{MEM},\mathscr{C}}$. For large enough n,

$$\hat{\nu}_n^{\mathrm{MEM},\mathscr{C}} = \frac{1}{n}\sum_{i=1}^{n} \psi'\left(\left\langle v_{n,\mathscr{C}}, \phi(x_i)\right\rangle\right)\delta_{x_i},$$

where $v_{n,\mathscr{C}}$ minimizes

$$H_n(v,\mathscr{C}) := \frac{1}{n}\sum_{i=1}^{n} \psi\left(\left\langle v, \phi(x_i)\right\rangle\right) - \inf_{Y\in\mathscr{C}} \langle v, Y\rangle.$$

We now have the following theorem.

THEOREM 2.1. *Suppose that Assumptions* $(H1)$, $(H2)$ *and* $(H3)$ *hold. Then any accumulation point* $\tilde{\nu}$ *of the sequence* $(\hat{\nu}_n^{\mathrm{MEM},\mathscr{C}})$ *has the form*

$$\tilde{\nu} = \psi'\left(\left\langle v_{\mathscr{C}}^*, \phi(x)\right\rangle\right)P + \sigma_{\mathscr{C}},$$

where $v_{\mathscr{C}}^*$ *is the unique minimum of the function*

$$H(v,\mathscr{C}) := \int_U \psi\left(\left\langle v, \phi(x)\right\rangle\right) dP - \inf_{Y\in\mathscr{C}} \langle v, Y\rangle$$

and $\sigma_{\mathscr{C}}$ *is a measure lying in* $\mathscr{R}_{v_{\mathscr{C}}^*} := \mathscr{S}(\tilde{\mathscr{C}})$, *with*

$$\tilde{\mathscr{C}} := \left\{Y - \int_U \phi(x)\psi'\left(\left\langle v_{\mathscr{C}}^*, \phi(x)\right\rangle\right) dP : Y \in \mathscr{C}\right\}.$$

Moreover, any element of $\mathscr{R}_{v_{\mathscr{C}}^*}$ *is concentrated on* $\Omega_\alpha(v_{\mathscr{C}}^*)$.

Theorem 2.1 proves that the MEM reconstruction may have singular parts. This was observed in a work of Livesey and Skilling (1985). Our result gives the theoretical background to their example:

Let $U := [0, 2\pi)^3$, $\phi_j(x^1, x^2, x^3) = \cos x^{j-1}$, $j = 2, 3, 4$. Let P be the uniform probability. The point $Y = (1, \xi, \xi, \xi)$ lies in the set of moments $\mathscr{K}$ (which will be precisely defined in Section 2.4) if and only if $-1 \le \xi \le 1$. Set $\mathscr{C} = \{Y\}$. Then taking for F the exponential distribution of parameter 1, we find $\psi'(\tau) = (1-\tau)^{-1}$, $\tau < \alpha = 1$. Now there is a critical value $\xi_0 \in (0,1)$ such that for any $\xi \in (0, \xi_0]$, the only accumulation point of the sequence

$(\hat{\nu}_n^{\mathrm{MEM},\mathscr{C}})$ is $\tilde{\nu} = (1 - \langle v_{\mathscr{C}}^*, \phi(x)\rangle)^{-1}P$, whereas if $\xi \in (\xi_0, 1)$, then $\tilde{\nu} = (1 - \langle v_{\mathscr{C}}^*, \phi(x)\rangle)^{-1}P + \lambda\delta_0$ for some positive real λ.

The MEM reconstruction reduces to the absolute continuous part in several cases.

CASE 1. If $\alpha = \infty$; see Dacunha-Castelle and Gamboa (1990) or Gamboa and Gassiat (1991).

CASE 2. If for any boundary point u of $D(F, \phi)$ we have

$$\lim_{\substack{v \to u \\ v \in D(F,\phi)}} \left\| \int_U \psi'(\langle v, \phi(x)\rangle)\phi(x)\, dP(x) \right\| = +\infty.$$

CASE 3. When $v_{\mathscr{C}}^*$ lies in the interior of $D(F, \phi)$.

We are able to give the dual characterization of the accumulation points (in the sense of convex analysis). For this purpose, let us introduce the function linked to H by the Legendre duality:

$$h_{\mathscr{C}}^*(c) := \sup_{v \in \mathbb{R}^k} (\langle v, c\rangle - H(v, \mathscr{C})), \qquad c \in \mathbb{R}^k.$$

THEOREM 2.2. *Under the same assumptions as in Theorem* 2.1,

$$h_{\mathscr{C}}^*(0) = \inf_{f \in \tilde{\mathscr{S}}(\mathscr{C})} \Gamma(f) = \min_{\mu \in \mathscr{S}(\mathscr{C})} J_F(\mu) = J_F(\tilde{\nu}),$$

where

$$\tilde{\mathscr{S}}(\mathscr{C}) := \{ f \in C(U) : fP \in \mathscr{S}(\mathscr{C})\} \tag{2.10}$$

and $\tilde{\nu}$ *is any accumulation point of* $(\hat{\nu}_n^{\mathrm{MEM},\mathscr{C}})$.

Theorem 2.2 gives a stochastic interpretation to a large class of deterministic convex methods. Indeed, if we solve the inverse problem by minimizing the convex functional J_F under the constraint

$$\int \phi\, d\mu \in \mathscr{C},$$

then the set of all the solutions contains all the accumulation points of the sequence of discretized estimates obtained by MEM.

Theorem 2.1 together with Theorem 2.2 proves that the reconstruction obtained by a deterministic convex method may have a singular part. A similar result was obtained after a preliminary version of this work by Borwein and Lewis (1993).

2.3. *The convergence theorem for* $\hat{\nu}_n^{\mathrm{bay},\mathscr{C}}$. The following theorem states the equivalence between MEM and the infinite-dimensional Bayesian estimator defined as the limit of the finite-dimensional Bayesian estimators.

THEOREM 2.3. *Assume that Assumptions* $(H1)$, $(H2)$ *and* $(H3)$ *hold. Assume that* $\mathscr{R}_{v^*_{\mathscr{C}}}$ *is either a singleton or empty. Then, as n tends to infinity, the sequence* $(\hat{\nu}_n^{\mathrm{bay},\mathscr{C}})$ *converges weakly to* $\hat{\nu}_\infty^{\mathrm{MEM},\mathscr{C}}$, *where* $\hat{\nu}_\infty^{\mathrm{MEM},\mathscr{C}}$ *is the unique accumulation point* (*and thus the limit*) *of the sequence* $(\hat{\nu}_n^{\mathrm{MEM},\mathscr{C}})$.

REMARKS. (i) The assumption on $\mathscr{R}_{v^*_{\mathscr{C}}}$ is always true when $v^*_{\mathscr{C}}$ lies in $D'(F, \phi)$.

(ii) In general, the usual convex methods [see Borwein and Lewis (1993)] have the general form (J_F is the convex criterion)

$$J_F(\mu) = \Gamma\left(\frac{d\mu}{dP}\right) + \alpha\left(\mu - \frac{d\mu}{dP}P\right)(U).$$

When Γ and α are associated with a distribution F as defined in Section 2.1, $\hat{\nu}_\infty^{\mathrm{MEM},\mathscr{C}}$ is the measure that minimizes J_F under the constraint $\int\phi\, d\nu \in \mathscr{C}$. Theorem 2.2 together with Theorem 2.3 says that the convex method is an asymptotic Bayesian method with prior $F^{\otimes n}$ as described previously.

(iii) Our Bayesian Theorem 2.3 is a direct consequence of the large deviations principle for $(\nu_n)_{n\in\mathbb{N}}$ proved in Section 3. Analogous results are obtained by Csiszar (1984, 1985), Van Campenhout and Cover (1981) and Robert (1990a, b), for i.i.d. variables: their results are a direct consequence of the usual Sanov theorem.

2.4. *The existence problem.* In this section we will assume that, in (1.1), μ is a probability measure ($Y_1 = 1$). The interesting statistical inverse problems occur when the parameter of interest μ could be recovered in the absence of noise through the observation Y. This situation can occur only near the boundary of $\mathscr{K}$, where $\mathscr{K}$ is the set of all ϕ-moments of probability distributions on U:

$$\mathscr{K} := \left\{Y \in \mathbb{R}^k : \exists\mu \in \mathscr{P}(U),\ \int_U \phi(x)\, d\mu(x) = Y\right\}.$$

Indeed, the boundary points are those around which superresolution phenomena can appear as described in Gassiat (1991), Gamboa and Gassiat (1994, 1996a) and Doukhan and Gamboa (1996). Roughly speaking, superresolution means that nonlinear methods (such as maximum entropy or L_1-norm minimization) taking into account the positivity information are drastically better than linear methods such as least squares. Points of the boundary are moments of singular measures with respect to P, in some situations purely discrete measures [see Krein and Nudel'man (1977)]. Intuitively, to be able to observe them via ν_n, we have to allow zero values and very large values for the Z_i at the same time; this is why we set $F(\{0\}) > 0$ and $\alpha < \infty$.

Now, around a boundary point and in the presence of noise, it could happen that the observation Y contradicts the prior information; that is, there exists no probability measure having Y as ϕ-moments. It is then quite important to be able to decide if, given an observation, there exist solutions to the inverse problem with noise level 0. In other words, to decide if Y lies in $\mathscr{K}$ or not. In this section, we give a family of criteria to answer this question.

To obtain our result, we need to strengthen Assumption (H2).

ASSUMPTION (H2′). For all $v \in \mathbb{R}^k \setminus \{0\}$, $P(\Omega_0(v)) = 0$.

Denote by L the function $L(Y) := h^*_{\{Y\}}(0) = -\inf_{v \in \mathbb{R}^k} H(v, \{Y\})$.

THEOREM 2.4. *Assume that Assumptions* $(H1)$ *and* $(H2')$ *hold and* $F(\{0\}) > 0$. *Then, for any* Y *in* $\mathbb{R}^k \cap \{Y \in \mathbb{R}^k,\ Y_1 = 1\}$:

(i) *The function* $L(Y) < -\log F(\{0\}) + \alpha$ *if and only if* Y *is in the interior of* $\mathscr{K}$.

(ii) *The function* $L(Y) = -\log F(\{0\}) + \alpha$ *if and only if* Y *is on the boundary of* $\mathscr{K}$.

(iii) *The function* $L(Y) = +\infty$ *if and only if* Y *is outside* $\mathscr{K}$.

Using considerations on convex sets, an existence criterion was given a long time ago [see, e.g., Krein and Nudel'man (1977), Theorem 1.1, page 58]:

$$Y \in \mathscr{K} \quad \Leftrightarrow \quad \forall v \in \mathscr{L}_+, \qquad \langle v, Y \rangle \geq 0, \tag{2.11}$$

where

$$\mathscr{L}_+ := \{ v \in \mathbb{R}^k : \forall x \in U, \langle v, \phi(x) \rangle \geq 0 \}. \tag{2.12}$$

This classical characterization can be viewed as an extremal version of Theorem 2.4. Theorem 2.4 says that Y lies in $\mathscr{K}$ if and only if, for all log-Laplace transforms of probabilities on $\mathbb{R}_+$ satisfying Assumption (H1) and $F(\{0\}) > 0$, we have

$$\forall v \in \mathscr{L}_+, \qquad -\langle v, Y \rangle + \alpha - \int_U \psi(-\langle v, \phi(x) \rangle + \alpha)\, dP(x) \leq -\log F(\{0\}) + \alpha,$$

that is,

$$\forall v \in \mathscr{L}_+, \qquad -\int_U \psi(-\langle v, \phi(x) \rangle + \alpha)\, dP(x) + \log F(\{0\}) \leq \langle v, Y \rangle.$$

Taking the supremum of this last equation over all F satisfying Assumption (H1) and $F(\{0\}) > 0$ leads to the criterion (2.11).

The functions L appear also as a new theoretical tool: see Gamboa and Gassiat (1996a) for applications in signal processing and Gamboa and Gassiat (1996b) for statistical applications. Similar ideas are developed for processes in Cattiaux and Léonard (1994) and for marginal problems in Cattiaux and Gamboa (1995).

EXAMPLE 1. If $U = [0, 2\pi)$, P is the uniform probability, $\phi_2(x) = \cos x$, $Y = (1, d)$ and F is the Poissonized exponential distribution (see example 6 in the Introduction; here $\beta = 1$), direct calculations lead to $L(Y) = 2[1 - (1 - d^2)^{1/4}]$ if $|d| \leq 1$ and $L(Y) = +\infty$ otherwise.

EXAMPLE 2. For $F := ((e-1)/e)\sum_{n=0}^{\infty} \exp(-n)\delta_n$, where δ_n is the Dirac mass at point n, the equation $L(Y) \leq -\log F(\{0\}) + \alpha$ is equivalent to

$$\forall v \in \mathbb{R}^k, \qquad \langle v, Y\rangle + \int_U \log\{1 - \exp\langle v, \phi(x)\rangle\}\, dP(x) \leq 0,$$

where $\log x := -\infty$ whenever $x \leq 0$.

2.5. *A posteriori asymptotic properties of* ν_n. Here, the omission of the subscript or exponent $\mathscr{C}$ means that $\mathscr{C}$ reduces to a singleton $\{Y\}$. We will now give some asymptotic properties of the random measure ν_n when $\mathbb{R}_+^n$ is endowed with P_n^{MEM}. The aim of the results is to state precisely the convergence of the discretized estimator to the infinite-dimensional reconstruction. We give two different results concerning the case where the limit is well defined, that is, where the set of accumulation points reduces to a singleton.

When v^* lies in $D'(F, \phi)$ and if $\mathbb{R}_+^n$ is endowed with P_n^{MEM}, we show the exponential convergence (in probability) of ν_n to $\hat{\nu}^{\mathrm{MEM}}$. Indeed, we have a large deviations principle for $(\nu_n)_{n\in\mathbb{N}}$ as soon as v^* lies in $D'(F, \phi)$.

PROPOSITION 2.5. *Let A be a measurable set of measures. If v^* lies in $D'(F,\phi)$, then*

$$\begin{aligned} -\Lambda_Y(\operatorname{int}(A)) &\leq \liminf_{n\to\infty} \frac{1}{n} \log P^{\mathrm{MEM}}_{n(\nu_n \in A)} \\ &\leq \limsup_{n\to\infty} \frac{1}{n} \log P^{\mathrm{MEM}}_{n(\nu_n \in A)} \\ &\leq -\Lambda_Y(\bar{A}), \end{aligned}$$

where $\Lambda_Y(A) := \inf_{\mu\in A} J_F^Y(\mu)$ *with*

$$J_F^Y(\mu) := J_F(\mu) - \Gamma(\psi'(\langle v^*, \phi(x)\rangle)) - \int_U \langle v^*, \phi(x)\rangle\, d\mu(x) + \langle v^*, Y\rangle.$$

When v^* does not lie in $D'(F, \phi)$ and $\mathscr{R}_{v^*}$ reduces to a singleton there exist situations where the distribution of ν_n (under P_n^{MEM}) does not converge to a delta distribution: there is no law of large numbers, though the expectation converges.

3. Proofs.

3.1. *Proof of Theorem* 2.1. We first show that if the sequence $(v_{n,\mathscr{C}})$ converges to $v^*_{\mathscr{C}}$, then the result holds. U and $\mathscr{C}$ are compact; thus the sequence $(\hat{\nu}_n^{\mathrm{MEM},\mathscr{C}})$ is tight. Let $\tilde{\nu}$ be an accumulation point of the sequence $(\hat{\nu}_n^{\mathrm{MEM},\mathscr{C}})$ and let (n_k) be an increasing sequence of integers such that $(\hat{\nu}_{n_k}^{\mathrm{MEM},\mathscr{C}})$ converges weakly, as k tends to infinity, to $\tilde{\nu}$. For any measure ν on U set

$$\mathscr{A}_\nu := \{A \in \mathscr{B}(U) : \nu(\partial A) = 0\},$$

where ∂A denotes the boundary of A. Let $A \in \mathscr{A}_P \cap \mathscr{A}_{\tilde{\nu}}$ such that there exists a closed subset F of U with $A \subset F$ and $F \cap \Omega_\alpha(v^*_{\mathscr{C}}) = \varnothing$. Then

$$\nu_{n_k}^{\mathrm{MEM},\mathscr{C}}(A) = \int_A \psi'(\langle v^*_{\mathscr{C}}, \phi(x)\rangle)\, dP + S_{1,n_k} + S_{2,n_k},$$

where

$$S_{1,n_k} := \frac{1}{n_k}\sum_{i=1}^{n_k}\Big(\psi'(\langle v_{n_i,\mathscr{C}}, \phi(x_i)\rangle) - \psi'(\langle v^*_{\mathscr{C}}, \phi(x_i)\rangle)\Big)1_A(x_i),$$

$$S_{2,n_k} := \frac{1}{n_k}\sum_{i=1}^{n_k}\psi'(\langle v^*_{\mathscr{C}}, \phi(x_i)\rangle)1_A(x_i) - \int_A \psi'(\langle v^*_{\mathscr{C}}, \phi(x)\rangle)\, dP.$$

However, because $A \subset F$ and F is closed and does not intersect $\Omega_\alpha(v^*_{\mathscr{C}})$, there exists a constant C such that for k large enough, $|S_{1,n_k}| \le C\|v_{n_k,\mathscr{C}} - v^*_{\mathscr{C}}\|$, so $\lim_{k\to\infty} S_{1,n_k} = 0$. As P_{n_k} converges weakly to P and $A \in \mathscr{A}_P$, we then get $\lim_{k\to\infty} S_{2,n_k} = 0$. Finally, since $A \in \mathscr{A}_{\tilde{\nu}}$,

$$\lim_{k\to\infty} \hat{\nu}_{n_k}^{\mathrm{MEM},\mathscr{C}}(A) = \tilde{\nu}(A) = \int_A \psi'(\langle v^*_{\mathscr{C}}, \phi(x)\rangle)\, dP.$$

Let $B \in \mathscr{A}_P \cap \mathscr{A}_{\tilde{\nu}}$ with $B \cap \Omega_\alpha(v^*_{\mathscr{C}}) = \varnothing$. For $\delta > 0$, set

$$A_\delta := \{x \in B : \langle v^*_{\mathscr{C}}, \phi(x)\rangle \le \alpha - \delta\}.$$

As δ decreases to 0, A_δ increases to B. Moreover, there exists a sequence (δ_n) decreasing to 0 with $\forall n \ge 1$, $A_{\delta_n} \in \mathscr{A}_P \cap \mathscr{A}_{\tilde{\nu}}$. Using the monotone convergence theorem, we then get

$$\tilde{\nu}(B) = \lim_{n\to\infty} \tilde{\nu}(A_{\delta_n}) = \lim_{n\to\infty}\int_{A_{\delta_n}} \psi'(\langle v^*_{\mathscr{C}}, \phi(x)\rangle)\, dP = \int_B \psi'(\langle v^*_{\mathscr{C}}, \phi(x)\rangle)\, dP.$$

Let now $B \in \mathscr{A}_P \cap \mathscr{A}_{\tilde{\nu}}$,

$$\begin{aligned}\tilde{\nu}(B) &= \tilde{\nu}(B \setminus (B \cap \Omega_\alpha(v^*_{\mathscr{C}}))) + \tilde{\nu}(B \cap \Omega_\alpha(v^*_{\mathscr{C}}))\\ &= \int_B \psi'(\langle v^*_{\mathscr{C}}, \phi(x)\rangle)\, dP + \sigma_{\mathscr{C}}(B),\end{aligned} \tag{3.1}$$

where $\sigma_{\mathscr{C}}$ is supported by $\Omega_\alpha(v^*_{\mathscr{C}})$. Now, since the σ-field generated by $\mathscr{A}_P \cap \mathscr{A}_{\tilde{\nu}}$ is $\mathscr{B}(U)$, (3.1) remains valid for any $B \in \mathscr{B}(U)$.

We shall now prove that the sequence $v_{n,\mathscr{C}}$ possesses a finite limit $v^*_{\mathscr{C}}$. Define for $c' \in \mathbb{R}^k$, $h^*_{n,\mathscr{C}}(c') := \sup_{v\in\mathbb{R}^k}(\langle v, c'\rangle - H_n(v,\mathscr{C}))$. As in Dacunha-Castelle and Gamboa (1990), using a separation theorem, it is not difficult to prove that for n sufficiently large, $h^*_{n,\mathscr{C}}$ is finite on a (nonvoid) open ball B centered in 0. Since $H_n(\cdot,\mathscr{C})$ is minimized at the point $v_{n,\mathscr{C}}$, 0 lies in $\partial H_n(v_{n,\mathscr{C}},\mathscr{C})$, the subdifferential of the function $H_n(\cdot,\mathscr{C})$ at the point $v_{n,\mathscr{C}}$. Using the fact that H_n is lower semicontinuous, we deduce from the Theorem 23.5 of Rockafellar [(1970), page 218] that $v_{n,\mathscr{C}} \in \partial h^*_{n,\mathscr{C}}(0)$.

Using Lemma 3.4 of Gamboa (1989), we see that the function $H(\cdot,\mathscr{C})$ attains its minimum at a unique point v^* and, using the same arguments as before, $\partial h^*_{\mathscr{C}}(0) = \{v^*\}$. Now using Theorem 24.5 of Rockafellar [(1970), page

233] on subdifferential sequence convergence, we get that $\lim_{n\to\infty} v_{n,\mathscr{C}} = v^*_{\mathscr{C}}$ as soon as $\lim_{n\to\infty} h^*_{n,\mathscr{C}}(c') = h^*_{\mathscr{C}}(c')$ for all c' lying in an open ball around 0. This is the aim of Proposition 3.1, which is proved in Section 3.2. □

PROPOSITION 3.1. *Under the assumptions of Theorem* 2.1,

$$\forall c' \in B, \qquad \lim_{n\to\infty} h^*_{n,\mathscr{C}}(c') = h^*_{\mathscr{C}}(c').$$

3.2. *Proof of Theorem* 2.2. We shall need the following technical result:

LEMMA 3.2. *If f_n is a nondecreasing sequence of strictly convex functions on $\mathbb{R}^k$ converging to f, if the minimum m_n of f_n is attained at x_n and if m, the minimum of f, is finite, then m_n converges to m.*

PROOF OF THEOREM 2.2. For any positive M define $\exp\psi_M(\tau)$, $\tau\in\mathbb{R}$, the Laplace transform of the truncated finite positive measure $F1_{[0,M]}$. Define as usual

$$\begin{aligned}
h^*_{M,\mathscr{C}}(c) &:= \sup_{v\in\mathbb{R}^k}\left\{\langle v,c\rangle + \inf_{Y\in\mathscr{C}}\langle v,Y\rangle - \int_U \psi_M(\langle v,\phi(x)\rangle)\,dP(x)\right\},\\
\gamma_M(t) &:= \sup_{\tau\in\mathbb{R}}\{\tau t - \psi_M(\tau)\},\\
\Gamma_M(f) &:= \int_U \gamma_M(f(x))\,dP(x).
\end{aligned}$$

Using Corollary 37.3.2 of Rockafellar [(1970), page 393],

$$h^*_{M,\mathscr{C}}(0) = \inf_{Y\in\mathscr{C}} h^*_{M,\{Y\}}(0)$$

and this infimum is attained at a point c^M in the interior of the cone generated by $\mathscr{K}$; that is, in the interior of

$$\mathscr{K}_\lambda := \{\lambda Y,\ \lambda\in\mathbb{R}_+,\ Y\in\mathscr{K}\}.$$

Now, if $c^M\in\operatorname{int}(\mathscr{K}_\lambda)$, there exists M_0 such that c^M is a ϕ-moment of a function bounded strictly by M_0 [apply, for example, results of Dacunha-Castelle and Gamboa (1990) or Gamboa and Gassiat (1991)].

Using the assumption on the support of F, it is easy to see that there exists a sequence of M converging to $+\infty$ such that $F1_{[0,M]}$ is a positive measure for which the convex hull of the support is $[0,M]$. It is now possible to apply the earlier results on MEM [we are in the case where $\alpha = +\infty$; see Gamboa and Gassiat (1991)] to obtain

$$\text{(3.2)}\quad \forall M > M_0, \qquad h^*_{M,\mathscr{C}}(0) < +\infty \quad\text{and}\quad h^*_{M,\mathscr{C}}(0) = \inf_{g\in\tilde{\mathscr{S}}(\mathscr{C})} \Gamma_M(g).$$

Define $f_M(v) = \int_U \psi_M(\langle v, \phi(x)\rangle)\, dP(x) - \inf_{Y \in \mathscr{C}} \langle v, Y\rangle$. Using Beppo–Levi's lemma, we have that the sequence f_M satisfies the assumptions of Lemma 3.2 with $f(\cdot) := H(\cdot, \mathscr{C})$ and then

$$h^*_{\mathscr{C}}(0) = \lim_{M \to +\infty} h^*_{M,\mathscr{C}}(0). \tag{3.3}$$

Using again Lemma 3.2, γ_M decreases to γ and Γ_M decreases to Γ. We have obviously $\inf_{f \in \tilde{\mathscr{S}}(\mathscr{C})} \Gamma_M(f) \geq \inf_{f \in \tilde{\mathscr{S}}(\mathscr{C})} \Gamma(f)$. Let ε be a positive number. Let g be a function of $C(U)$ such that $\Gamma(g) \leq \inf_{f \in \tilde{\mathscr{S}}(\mathscr{C})} \Gamma(f) + \varepsilon$. Then, for M large enough, $\Gamma_M(g) \leq \Gamma(g) + \varepsilon$. So $\inf_{f \in \tilde{\mathscr{S}}(\mathscr{C})} \Gamma_M(f) \leq \inf_{f \in \tilde{\mathscr{S}}(\mathscr{C})} \Gamma(f) + 2\varepsilon$. We may deduce that

$$\lim_{M \to \infty} \inf_{f \in \tilde{\mathscr{S}}(\mathscr{C})} \Gamma_M(f) = \inf_{f \in \tilde{\mathscr{S}}(\mathscr{C})} \Gamma(f). \tag{3.4}$$

From results (3.2), (3.3) and (3.4) we deduce the first equality of Theorem 2.2. The second and the third follow from Lemma 3.3. □

Proof of Proposition 3.1. Let B be an open ball that contains 0 and such that, for large enough n, $h^*_{n,\mathscr{C}}$ is finite on B. By definition, $h^*_{n,\mathscr{C}}(c') \geq -H_n(v, \mathscr{C}) + \langle v, c'\rangle$ for all c' in B and v in $\mathbb{R}^k$. Consequently,

$$\forall v \in D'(F, \phi), \qquad \liminf_{n \to \infty} h^*_{n,\mathscr{C}}(c') \geq -H(v, \mathscr{C}) + \langle v, c'\rangle.$$

Now H is lower semicontinuous and $D(F, \phi)$ is included in the closure of $D'(F, \phi)$. Therefore, $\inf_{v \in D'(F,\phi)}(H(v, \mathscr{C}) - \langle v, c'\rangle) = \inf_{v \in D(F,\phi)}(H(v, \mathscr{C}) - \langle v, c'\rangle)$. Then

$$\liminf_{n \to \infty} h^*_{n,\mathscr{C}}(c') \geq h^*_{\mathscr{C}}(c'). \tag{3.5}$$

Define c^0 by $h^*_{\mathscr{C}}(0) = \inf_{Y \in \mathscr{C}} h^*_{\{Y\}}(0) = h^*_{\{c^0\}}(0)$. We have

$$\forall c' \in B, \exists f \in C(U), \qquad \int_U f(x)\phi(x)\, dP(x) = c^0 + c'.$$

For all $M \geq 2\|f\|_\infty$ we can apply the construction with ψ_M (see the proof of Theorem 2.2), so that

$$h_{n,\mathscr{C}}(c') \leq (1/n) \sum_{i=1}^{n} \gamma_M\left[\psi'_M\left(\left\langle v^M_{n,\mathscr{C}}, \phi(x_i)\right\rangle\right)\right],$$

$$\limsup_{n \to \infty} h_{n,\mathscr{C}}(c') \leq \Gamma_M\left(\psi'_M\left(\left\langle v^{*M}_{\mathscr{C}}, \phi(x)\right\rangle\right)\right) = h^*_{m,\mathscr{C}}(c'),$$

where $v^M_{n,\mathscr{C}}$ and $v^*_{\mathscr{C}}$ are the minimizers defined in Section 2 and in Theorem 2.1 for MEM with the truncated measure $F1_{[0,M]}$. Taking the limit as M goes to infinity, we get

$$\limsup_{n \to \infty} h_{n,\mathscr{C}}(c') \leq h^*_{\mathscr{C}}(c'). \tag{3.6}$$

Equations (3.5) and (3.6) give Proposition 3.1. □

LEMMA 3.3. *Any accumulation point of the sequence* $\hat{\nu}_n^{\mathrm{MEM},\mathscr{C}}$ *is a minimizer of* J_F *on* $\mathscr{S}(\mathscr{C})$.

PROOF. Let $\tilde{\nu}$ be an accumulation point of the sequence $\hat{\nu}_n^{\mathrm{MEM},\mathscr{C}}$. Let $f \in C(U)$ be such that $\int_U \phi f \, dP \in \mathscr{C}$. Using the proof of Lemma 6 in Gamboa and Gassiat (1991), we have

$$\Gamma_n\big(\hat{\nu}_n^{\mathrm{MEM},\mathscr{C}}\big) \le \Gamma_n(f)$$

with

$$\Gamma_n\big(\hat{\nu}_n^{\mathrm{MEM},\mathscr{C}}\big) = \frac{1}{n}\sum_{i=1}^{n} \gamma\big[\psi'\big(\langle v_{n,\mathscr{C}}, \phi(x_i)\rangle\big)\big]$$

and

$$\Gamma_n(f) = \frac{1}{n}\sum_{i=1}^{n} \gamma(f(x_i)).$$

Using the same arguments as in Section 3.1, it is easy to show that

$$\lim_{n\to\infty} \Gamma_n\big(\hat{\nu}_n^{\mathrm{MEM},\mathscr{C}}\big) = J_F(\tilde{\nu}),$$
$$\lim_{n\to\infty} \Gamma_n(f) = \Gamma(f),$$

so that

$$J_F(\tilde{\nu}) \le \inf_{f \in \tilde{\mathscr{S}}(\mathscr{C})} \Gamma(f),$$

which, together with the second equality of Theorem 2.2, proves the lemma. $\square$

3.3. *Large deviations principle for* $(\nu_n)_{n\in\mathbb{N}}$. To prove the results of Section 2.3, we make extensive use of large deviation results. We give them here. Let $(z_n)_{n\in\mathbb{N}}$ be a random sequence. We will say in the sequel that $(z_n)_{n\in\mathbb{N}}$ obeys a large deviations principle if the sequence of its distributions obeys a large deviations principle. Let Q_n be the probability distribution on $\mathscr{M}_+(U)$ of ν_n [ν_n defined by (1.3)] when $\mathbb{R}_+^n$ is endowed with $F^{\otimes n}$. Then we have the following large deviations principle for $(\nu_n)_{n\in\mathbb{N}}$:

THEOREM 3.4. *For any subset B of* $\mathscr{M}(U)$, *let* $\Lambda(B) := \inf_{\mu\in B} J_F(\mu)$. *Assume that Assumption* $(H1)$ *holds. Then for every Borel subset A of* $\mathscr{M}(U)$,

$$\begin{aligned} -\Lambda(\operatorname{int}(A)) &\le \liminf_{n\to+\infty} \frac{1}{n}\log Q_n(A) \\ &\le \limsup_{n\to+\infty} \frac{1}{n}\log Q_n(A) \le -\Lambda(\bar{A}). \end{aligned} \tag{3.7}$$

This theorem is proved by applying the following abstract large deviation result. Let X be the dual space of a Banach space X'. We endow X with the weak-$*$ topology so that X′ is the dual of X. In this subsection, $\langle \cdot, \cdot \rangle$ will be the dual product between X′ and X. If Q is a probability measure on X, we define its log-Laplace transform as

$$\forall f \in X', \qquad H_Q(f) := \log \int_X \exp\langle f, x\rangle Q(dx).$$

Let now $(Q_h)_{h>0}$ be a family of probability distributions on X and let $H_h := H_{Q_h}$ be their log-Laplace transforms. We shall need the following assumptions.

ASSUMPTION (H4). There exists a function $l(h)$ from $\mathbb{R}_+$ to $\mathbb{R}_+$ and a function H from X' to $\mathbb{R} \cup \{+\infty\}$ such that H is convex, lower semicontinuous and, if $D(H)$ is the domain of H (that is, the set of points of X' where H is finite), $D(H)$ contains an open neighborhood of 0. Function H satisfies

$$\begin{aligned} &\forall f \in \operatorname{int}(D(H)), \qquad \lim_{h\to+\infty} \frac{1}{l(h)} H_h(l(h)f) = H(f), \\ &\qquad\qquad\qquad\qquad\qquad \lim_{h\to+\infty} l(h) = +\infty. \end{aligned} \tag{3.8}$$

ASSUMPTION (H5). For every $R > 0$, there exists a compact set Ξ_R of X such that

$$\limsup_{h\to+\infty} \frac{1}{l(h)} \log Q_h(\Xi_R^c) \le -R. \tag{3.9}$$

ASSUMPTION (H6). If f is a boundary point of $D(H)$, where $H(f)$ is finite, then there exists a sequence of points f_n in $\operatorname{int}(D(H))$ such that $H(f_n)$ converges to $H(f)$.

Observe that, because $D(H)$ contains an open set, H is continuous in $\operatorname{int}(D(H))$ [see Ekeland and Temam (1976), Proposition 2.5, page 12].

DEFINITION 1. $I(x) := \sup_{f\in X'}\{\langle f, x\rangle - H(f)\}$, $x \in X$, denotes the Legendre transform of H and $D(I)$ is its domain.

DEFINITION 2. For every set A included in X, we define $\Lambda(A) := \inf_{x\in A} I(x)$.

ASSUMPTION (H7). For every open set A such that $\Lambda(A)$ is finite and for all positive η, there exists x in $D(I) \cap A$ such that:

(i) $\partial I(x) = \{f\}$, with $f \in \operatorname{int}(D(H))$;
(ii) I is strictly convex at x;
(iii) $I(x) \le \Lambda(A) + \eta$.

Where $\partial I(x)$ denotes the subdifferential of I at point x [see Definition 5.1 of Ekeland and Temam (1976), page 20]. Under these assumptions, the sequence Q_h obeys a large deviations principle; that is, we have the following proposition:

PROPOSITION 3.5. *Suppose that Assumptions* $(H4)$, $(H5)$, $(H6)$ *and* $(H7)$ *hold. Then, for every Borel subset* A *of* X,

$$
\begin{aligned}
-\Lambda(\operatorname{int}(A)) &\le \liminf_{h\to+\infty} \frac{1}{l(h)} \log Q_h(A) \\
&\le \limsup_{h\to+\infty} \frac{1}{l(h)} \log Q_h(A) \le -\Lambda(\bar{A}).
\end{aligned} \tag{3.10}
$$

The proof of this proposition follows the proof of Theorem 1.1 of Baldi (1988), with minor modifications, due to the fact that our assumptions are weaker than those used by Baldi on the boundary of $D(H)$. We shall omit the details.

Let us now prove Theorem 3.4. Here, $X := \mathscr{M}(U)$, $X' := C(U)$, $h := n$, $l(h) := n$ and we have

$$
\begin{aligned}
\forall f \in C(U), \qquad H_n(f) &= E_{Q_n} \exp \int_U f(x)\, d\nu_n(x) \\
&= \exp \sum_{i=1}^{n} \psi\left(\frac{f(x_i)}{n}\right).
\end{aligned}
$$

Define, for all real continuous functions f on U, $H(f) = \int_U \psi(f(x))\, dP(x)$. Proof of the theorem proceeds by a verification of assumptions of Proposition 3.5 and will be omitted. We just give the explicit form of the Legendre transform of H. Let $\tilde{J}_F$ be the Legendre transform of H:

$$
\forall \mu \in \mathscr{M}(U), \qquad \tilde{J}_F(\mu) = \sup_{f \in C(U)} \left(\int_U f(x)\, d\mu(x) - \int_U \psi(f(x))\, dP(x) \right).
$$

We prove that $\tilde{J}_F$ coincides with the functional J_F defined in Section 2.1.

PROPOSITION 3.6. *If* μ *is a positive bounded measure on* U *with Radon-Nikodym decomposition* $\mu = gP + \sigma$, *then*

$$
\tilde{J}_F(\mu) = \begin{cases} \alpha\sigma(U) + \Gamma(g) = J_F(\mu), \\ +\infty, \quad \textit{otherwise}. \end{cases}
$$

PROOF. To compute the Legendre transform of H, we shall make use of Theorem 5 of Rockafellar (1971). Indeed its assumptions hold here: U is compact, $P(V) \neq 0$ provided that V is a nonempty open subset of U and (ii) $D(\psi) = (-\infty, \alpha)$. We may apply the conclusion

$$
\begin{aligned}
&\sup_{f \in C(U)} \left(\int_U f(x)\, d\mu(x) - \int_U \psi(f(x))\, dP(x) \right) \\
&\qquad = \int_U \gamma\left(\frac{d\mu}{dP}\right) dP + \int_U r\left(\frac{d\mu}{d\theta}\right) d\theta,
\end{aligned}
$$

where σ is the singular part of μ with respect to P, θ is any measure in $\mathscr{M}(U)$ such that $\sigma \ll \theta$ and r is the recession function [see Rockafellar (1970), page 66] of γ. Using Theorem 8.5 of Rockafellar [(1970), page 66], it is easy to see that

$$r(x) = \begin{cases} \alpha x, & \text{if } x > 0, \\ 0, & \text{if } x = 0, \\ +\infty, & \text{if } x < 0. \end{cases}$$

Now $\gamma(t) = +\infty$ for $t < 0$, so that if $\mu \notin \mathscr{M}^+(U)$, $\tilde{J}_F(\mu) = +\infty$. Moreover, if $\mu \in \mathscr{M}^+(U)$, then

$$\tilde{J}_F(\mu) = \int_U \gamma\left(\frac{d\mu}{dP}\right) dP + \int_U r\left(\frac{d\mu}{d\theta}\right) d\theta = \Gamma(g) + \alpha\sigma(U). \qquad \square$$

REMARKS. (i) The domain $D(J_F)$ is the set of positive measures for which the P-absolutely continuous part, say fP, satisfies $\Gamma(f) < +\infty$.

(ii) The function γ is strictly convex and the total mass of a measure is a linear function. The points where J_F is strictly convex are then obviously the P-absolutely continuous positive measures.

3.4. *Proof of Theorem* 2.3. First of all, we shall denote the conditioning set $\{\int \phi \, d\nu_n \in \mathscr{C}\}$ by $\mathscr{E}_n$. For $f \in C(U)$, we have

$$\left| \int_U f(x)\, d\hat{\nu}_n^{\mathrm{bay},\mathscr{C}}(x) - \int_U f(x)\, d\nu_\infty^{\mathrm{MEM},\mathscr{C}}(x) \right| = \left| E_{F^{\otimes n}}[W_n | \mathscr{E}_n] \right|$$

$$\leq \int_0^\infty F^{\otimes n}\{\{|W_n| > t\} | \mathscr{E}_n\}\, dt$$

with $W_n := \int_U f(x) d(\nu_n - \nu_\infty^{\mathrm{MEM}\,\mathscr{C}})(x)$. We will now show that $A_n := \int_0^\infty F^{\otimes n}\{\{|W_n| > t\} \,|\, \mathscr{E}_n\}\, dt$ tends to 0 when n tends to infinity.

STEP 1. Let

$$\Theta_n := \left\{ (z_1, \dots, z_n) \in \mathbb{R}_+^n : (1/n) \sum_{i=1}^n z_i \phi(x_i) \in \operatorname{int}(\mathscr{C}) \cap \mathscr{K}_\lambda \right\}.$$

Using Lemma 1 of Dacunha-Castelle and Gamboa (1990), we have that for n large enough $F^{\otimes n}(\Theta_n) > 0$. Then, for large enough n and for any $t > 0$, we have

$$F^{\otimes n}\{|W_n| > t | \mathscr{E}_n\} = \frac{F^{\otimes n}(\{|W_n| > t\} \cap \mathscr{E}_n)}{F^{\otimes n}(\mathscr{E}_n)}.$$

STEP 2. Apply now the large deviations principle for $(\nu_n)_{n \in \mathbb{N}}$ (Theorem 3.4):

$$\liminf_{n\to\infty} \frac{1}{n} \log F^{\otimes n}(\mathscr{E}_n) \geq -\Lambda(\operatorname{int}(\mathscr{S}(\mathscr{C})))$$

and

$$\limsup_{n\to\infty} \frac{1}{n} \log F^{\otimes n}(\{|W_n| > t\} \cap \mathscr{E}_n) \leq -\Lambda(B_t \cap \mathscr{S}(\mathscr{C}))$$

with

$$B_t := \left\{ \nu \in \mathscr{M}^+(U) : \left| \int_U f(x)\, d\left(\nu - \nu_\infty^{\mathrm{MEM},\mathscr{C}}\right)(x) \right| \geq t \right\}.$$

Using the semicontinuity of J_F, it is easy to see that $\Lambda(\mathrm{int}(\mathscr{S}(\mathscr{C}))) = \Lambda(\mathscr{S}(\mathscr{C}))$, and Theorem 2.2 gives $\Lambda(\mathscr{S}(\mathscr{C})) = J_F(\nu_\infty^{\mathrm{MEM},\mathscr{C}})$. Moreover, because $\nu_\infty^{\mathrm{MEM},\mathscr{C}}$ is the unique minimum point of J_F on $\mathscr{S}(\mathscr{C})$, we have, for any positive real t, $J_F(\nu_\infty^{\mathrm{MEM},\mathscr{C}}) < \Lambda(B_t \cap \mathscr{S}(\mathscr{C}))$. Thus, for any positive t, there exists $\varepsilon > 0$ such that

$$J_F\left(\nu_\infty^{\mathrm{MEM},\mathscr{C}}\right) < -2\varepsilon + \Lambda\left(B_t \cap \mathscr{S}(\mathscr{C})\right). \tag{3.11}$$

Now, for this choice of ε, $\exists n_1$, $\forall n \geq n_1$,

$$\log F^{\otimes n}\left\{\{|W_n| > t\} \middle| \mathscr{E}_n\right\} \leq nJ_F\left(\nu_\infty^{\mathrm{MEM},\mathscr{C}}\right) + n\varepsilon - n\Lambda\left(B_t \cap \mathscr{S}(\mathscr{C})\right).$$

As n goes to infinity, using (3.11) we finally get $\lim_{n\to\infty} F^{\otimes n}\{\{|W_n| > t\} \mid \mathscr{E}_n\} = 0$. The Lebesgue convergence theorem allows us now to conclude that $\lim_{n\to\infty} A_n = 0$. $\square$

3.5. *Proof of Theorem* 2.4. Applying Theorem 3.4 and the contraction principle [Remark 1 in Varadhan (1984), page 5], the large deviations functional for $(\int_U \phi(x)\, d\nu_n(x))_{n\in\mathbb{N}}$ is $\inf_{\mu\in\mathscr{S}(\{Y\})} J_F(\mu)$. On the other hand, from Theorem 2.2 and Theorem 3 of Rockafellar (1971), we have that $\inf_{\mu\in\mathscr{S}(\{Y\})} J_F(\mu) = L(Y)$, where

$$L(Y) := \sup_{v\in\mathbb{R}^k} \left[\langle v, Y \rangle - \int_U \psi(\langle v, \phi(x) \rangle)\, dP(x) \right], \qquad Y \in \mathbb{R}^k.$$

The last assertion of the theorem is straightforward when we note that if Y (with $Y_1 = 1$) is outside of $\mathscr{K}$, any μ that satisfies $\int_U \phi(x)\, d\mu(x) = Y$ is signed. To prove the first and second assertions, we first study the behavior of γ. Let $m := \int_{\mathbb{R}_+} y\, dF(y) = \psi'(0)$. As γ is a convex function and $\gamma(0) = -\log F(\{0\})$, $\gamma(m) = \gamma(m) = 0$, we have

$$\forall y \in [0, m], \qquad \gamma(y) \leq -\log F(\{0\}), \tag{3.12}$$

$$\forall y \in [m, +\infty), \qquad \gamma(y) \leq \alpha(y - m) \leq \alpha y. \tag{3.13}$$

If Y is in $\mathscr{K}$, there exists a probability measure ν on U such that $\int_U \phi(x)\, d\nu(x) = Y$ and $L(Y) \leq J_F(\nu)$. Now

$$\begin{aligned} J_F(\nu) = &\int_{d\nu/dP \leq m} \gamma\left(\frac{d\nu}{dP}(x)\right) dP(x) \\ &+ \int_{d\nu/dP > m} \gamma\left(\frac{d\nu}{dP}(x)\right) dP(x) + \alpha\left(\nu - \frac{d\nu}{dP}P\right)(U), \end{aligned}$$

so using (3.12) and (3.13),

$$
\begin{aligned}
J_F(\nu) \le &-\log F(\{0\}) \int_{d\nu/dP \le m} dP(x) \\
&+ \alpha \int_{d\nu/dP \ge m} \frac{d\nu}{dP}\, dP(x) + \alpha\left(\nu - \frac{d\nu}{dP} P\right)(U),
\end{aligned}
$$

so that $L(Y) \le -\log F(\{0\}) + \alpha$.

If Y is in the interior of $\mathscr{K}$, ν can be chosen absolutely continuous with respect to P and with positive continuous density. Hence, the inequality is strict.

Finally, if Y is on the boundary of $\mathscr{K}$, any ν such that $\int_U \phi(x)\, d\nu(x) = Y$ satisfies $d\nu/dP = 0$ [ν is supported by a level set $\Omega_0(\xi)$, where $\xi \in \mathscr{L}_+$ satisfies $\langle \xi, Y\rangle = 0$], and the inequality becomes equality. □

3.6. *Proof of Proposition* 2.5. We can directly apply Theorem 3.5. The proof follows the same line as the proof of Proposition 3.4. We just have to calculate the large deviations functional J_F^Y. Let f be an element of $C(U)$ such that $\forall x \in U$, $f(x) + \langle v', \phi(x)\rangle < \alpha$. Then

$$
\begin{aligned}
&\lim_{n\to\infty} \frac{1}{n} \log E_{P_n^{\mathrm{MEM}}} \exp\left(n \int_U f(x)\, d\nu_n(x)\right) \\
&\quad = \lim_{n\to\infty} \frac{1}{n} \log E_{F^{\otimes n}} \exp\Bigg(n \int_U \big(f(x) + \langle v_n, \phi(x)\rangle\big)\, d\nu_n(x) \\
&\qquad\qquad - \sum_{i=1}^{n} \psi\big(\langle v_n, \phi(x_i)\rangle\big)\Bigg) \\
&\quad = \lim_{n\to\infty} \left(\frac{1}{n} \sum_{i=1}^{n} \psi\big(f(x_i) + \langle v_n, \phi(x_i)\rangle\big) - \frac{1}{n} \sum_{i=1}^{n} \psi\big(\langle v_n, \phi(x_i)\rangle\big)\right) \\
&\quad = \int_U \big(\psi(f(x) + \langle v^*, \phi(x)\rangle) - \psi(\langle v^*, \phi(x)\rangle)\big)\, dP(x).
\end{aligned}
$$

On the other hand, if f is such that $\exists x \in U$, $f(x) + \langle v^*, \phi(x)\rangle > \alpha$, then

$$
\lim_{n\to\infty} \frac{1}{n} \log E_{P_n^{\mathrm{MEM}}} \exp\left(n \int_U f(x)\, d\nu_n(x)\right) = \infty.
$$

Consequently,

$$
\begin{aligned}
J_F^Y(\mu) = \sup_{f \in C(U)} \Bigg\{ &\int_U f(x)\, d\mu(x) \\
&- \int_U \big[\psi(f(x) + \langle v^*, \phi(x)\rangle) - \psi(\langle v^*, \phi(x)\rangle)\big]\, dP(x) \Bigg\}
\end{aligned}
$$

$$= \sup_{h \in C(U)} \left\{ \int_U h(x)\, d\mu - \int_U \psi(h(x))\, dP(x) \right\}$$

$$+ \int_U \psi(\langle v^*, \phi(x) \rangle)\, dP(x) - \int_U \langle v^*, \phi(x) \rangle\, d\mu(x)$$

$$= J_F(\mu) - \Gamma(\psi'(\langle v^*, \phi(x) \rangle)) - \int_U \langle v^*, \phi(x) \rangle\, d\mu(x) + \langle v^*, Y \rangle. \quad \square$$

Acknowledgments. The authors want to thank P. Cattiaux, F. Comets, D. Dacunha-Castelle and C. Léonard for helpful discussions.

REFERENCES

BALDI, P. (1988). Large deviations and stochastic homogenization. *Ann. Mat. Pura Appl.* **11** 161–177.

BARRY, D. (1986). Nonparametric Bayesian regression. *Ann. Statist.* **14** 934–953.

BORWEIN, J. M. and LEWIS, A. S. (1993). Partially-finite programming in L_1 and the existence of maximum entropy estimates. *SIAM J. Optim.* **3** 248–267.

CATTIAUX, P. and LÉONARD, C. (1994). Minimization of Kullback information for diffusion processes. *Ann. Inst. H. Poincaré Probab. Statist.* **30** 83–132.

CATTIAUX, P. and GAMBOA, F. (1995). Large deviations and variational theorems for marginal problems. Preprint, Orsay.

CSISZAR, I. (1984). Sanov property, generalized *I*-projection and a conditional limit theorem. *Ann. Probab.* **12** 768–793.

CSISZAR, I. (1985). An extended maximum entropy principle and a Bayesian justification theorem. In *Bayesian Statistics* **2** 83–98. North-Holland, Amsterdam.

CSISZAR, I. (1991). Why least squares and maximum entropy? An axiomatic approach to inference for linear inverse problems. *Ann. Statist.* **19** 2032–2066.

DACUNHA-CASTELLE, D. and DUFLO, M. (1986). *Probability and Statistics*. Springer, New York.

DACUNHA-CASTELLE, D. and GAMBOA, F. (1990). Maximum d'entropie et problème des moments. *Ann. Inst. H. Poincaré Probab. Statist.* **26** 567–596.

DIACONIS, P. and FREEDMAN, D. (1986). On the consistency of Bayes estimates. *Ann. Statist.* **14** 1–26.

DONOHO, D., JOHNSTONE, I., HOCH, J. and STERN, A. (1992). Maximum entropy and the nearly black object. *J. Roy. Statist. Soc. Ser. B* **54** 41–82.

DONOHO, D. (1993). Superresolution via sparsity constraints. *SIAM J. Math. Anal.* **23** 1309–1331.

DONOHO, D. and GASSIAT, E. (1992). Superresolution via positivity constraints. Unpublished manuscript.

DOUKHAN, P. and GAMBOA, F. (1996). Superresolution rates in Prokhorov metric. *Canad. J. Math.* **48** 316–329.

EKELAND, I. and TEMAM, R. (1976). *Convex Analysis and Variational Problems*. North-Holland, Amsterdam.

FERGUSON, T. S. (1974). Prior distributions on spaces of probability. *Ann. Statist.* **2** 615–629.

FRIEDEN, B. R. (1985). Dice, entropy and likelihood. *Proc. IEEE* **73** 1764–1770.

GAMBOA, F. (1989). Méthode du maximum d'entropie sur la moyenne et applications. Thèse Orsay.

GAMBOA, F. (1994). New Bayesian methods for ill posed problems. Preprint, Orsay.

GAMBOA, F. and GASSIAT, E. (1991). Maximum d'entropie et problème des moments: cas multidimensionnel. *Probab. Math. Statist.* **12** 67–83.

GAMBOA, F. and GASSIAT, E. (1994). The maximum entropy method on the mean: applications to linear programming and superresolution. *Math. Programming Ser. A* **66** 103–122.

GAMBOA, F. and GASSIAT, E. (1996a). Sets of superresolution and the maximum entropy method on the mean. *SIAM J. Math. Anal.* **27** 1129–1152.

GAMBOA, F. and GASSIAT, E. (1996b). Blind deconvolution of discrete linear systems. *Ann. Statist.* **24** 1964–1981.

GASSIAT, E. (1991). Problème des moments et concentration de mesures *C. R. Acad. Sci. Paris Sér. I Math.* **310** 41–44.

KREIN, M. G. and NUDEL'MAN, A. A. (1977). The Markov moment problem and extremal problems. *Translations of Mathematical Monographs* **50**. Amer. Math. Soc., Providence, RI.

LIVESEY, A. K. and SKILLING, J. (1985). Maximum entropy theory. *Acta Cryst. Sect. A* **41** 113–122.

MCLAUGHLIN, D. W. (1984). *Inverse Problems*. Amer. Math. Soc., Providence, RI.

ROBERT, C. (1990a). An entropy concentration theorem. *J. Appl. Probab.* **27** 303–313.

ROBERT, R. (1990b). Etats d'équilibre statistique pour l'écoulement bidimensionnel d'un fluide parfait. *C. R. Acad. Sci. Paris Sér. I Math.* **311** 575–578.

ROCKAFELLAR, R. T. (1970). *Convex Analysis*. Princeton Univ. Press.

ROCKAFELLAR, R. T. (1971). Integrals which are convex functionals II. *Pacific J. Math.* **39** 439–469.

VAN CAMPENHOUT, J. M. and COVER, T. M. (1981). Maximum entropy and conditional probability. *IEEE Trans. Inform. Theory* **27** 483–489.

VARADHAN, S. R. S. (1984). *Large Deviations and Applications*. SIAM, Philadelphia, PA.

WAHBA, G. (1990). *Spline Models for Observational Data*. SIAM, Philadelphia, PA.

LABORATOIRE DE STATISTIQUES
UNIVERSITÉ PARIS SUD
91405 ORSAY
FRANCE

LABORATOIRE ANALYSE ET PROBABILITÉ
UNIVERSITÉ D'EVRY-VAL D'ESSONNE
91425 EVRY
FRANCE