Regularization with Approximated L^2 Maximum Entropy Method

Jean-Michel Loubes, Paul Rochet

▶ **To cite this version:**

HAL Id: hal-00389698

https://hal.archives-ouvertes.fr/hal-00389698

Submitted on 2 Jun 2009

Regularization with Approximated L^2 Maximum Entropy Method

J-M. Loubes and P. Rochet

Abstract We tackle the inverse problem of reconstructing an unknown finite measure μ from a noisy observation of a generalized moment of μ defined as the integral of a continuous and bounded operator Φ with respect to μ. When only a quadratic approximation Φ_m of the operator is known, we introduce the L^2 approximate maximum entropy solution as a minimizer of a convex functional subject to a sequence of convex constraints. Under several assumptions on the convex functional, the convergence of the approximate solution is established and rates of convergence are provided.

1 Introduction

A number of inverse problems may be stated in the form of reconstructing an unknown measure μ from observations of generalized moments of μ, i.e., moments y of the form

$$y = \int_{\mathscr{X}} \Phi(x) d\mu(x),$$

where $\Phi : \mathscr{X} \to \mathbb{R}^k$ is a given map. Such problems are encountered in various fields of sciences, like medical imaging, time-series analysis, speech processing, image restoration from a blurred version of the image, spectroscopy, geophysical sciences, crytallography, and tomography; see for example Decarreau et al (1992), Gzyl (2002), Hermann and Noll (2000), and Skilling (1988). Recovering the unknown measure μ is generally an ill-posed problem, which turns out to be difficult

J-M. Loubes

Institut de Mathématiques de Toulouse, UMR 5219, Université Toulouse 3, 118 route de Narbonne F-31062 Toulouse Cedex 9, France e-mail: `loubes@math.univ-toulouse.fr`

P. rochet

Institut de Mathématiques de Toulouse, UMR 5219, Université Toulouse 3, 118 route de Narbonne F-31062 Toulouse Cedex 9, France e-mail: `rochet@math.univ-toulouse.fr`

to solve in the presence of noise, i.e., one observes y^{obs} given by

$$y^{obs} = \int_{\mathscr{X}} \Phi(x) d\mu(x) + \varepsilon. \tag{1}$$

For inverse problems with known operator Φ, regularization techniques allow the solution to be stabilized by giving favor to those solutions which minimize a regularizing functional J, i.e., one minimizes $J(\mu)$ over μ subject to the constraint that $\int_{\mathscr{X}} \Phi(x) d\mu(x) = y$ when y is observed, or $\int_{\mathscr{X}} \Phi(x) d\mu(x) \in K_Y$ in the presence of noise, for some convex set K_Y containing y^{obs}. Several types of regularizing functionals have been introduced in the literature. In this general setting, the inversion procedure is deterministic, i.e., the noise distribution is not used in the definition of the regularized solution. Bayesian approaches to inverse problem allow one to handle the noise distribution, provided it is known, yet in general, a distribution like the normal distribution is postulated (see Evans and Stark, 2002 for a survey). However in many real-world inverse problems, the noise distribution is unknown, and only the output y is easily observable, contrary to the input to the operator. Consequently very few paired data is available to reliably estimate the noise distribution, thereby causing robustness deficiencies on the retrieved parameters. Nonetheless, even if the noise distribution is unavailabe to the practitioner, she often knows the *noise level*, i.e., the maximal magnitude of the disturbance term, say $\rho > 0$, and this information may be reflected by taking a constraint set K_Y of diameter 2ρ.

As an alternative to standard regularizations such as Tikhonov or Galerkin, see for instance Engl, Hanke and Neubauer (1996), we focus on a regularization functional with grounding in information theory, generally expressed as a negative entropy, leading to *maximum entropy* solutions to the inverse problem. In a deterministic framework, maximum entropy solutions have been studied in Borwein and Lewis (1993, 1996), while some others study exist in a Bayesian setting (Gamboa, 1999; Gamboa and Gassiat, 1997), in seismic tomography (Fermin, Loubes and Ludeña, 2006), in image analysis (Gzyl and Zeev, 2003; Skilling and Gull, 2001). Regularization with maximum entropy also provides one with a very simple and natural manner to incorporate constraints on the support and the range of the solution (see e.g. the discussion in Gamboa and Gassiat, 1997).

In many actual situations, however, the map Φ is unknown and only an approximation to it is available, say Φ_m, which converges in quadratic norm to Φ as m goes to infinity. In this paper, following lines devised in Gamboa (1999) and Gamboa and Gassiat (1999) and Loubes and Pelletier (2008), we introduce an approximate maximum entropy on the mean (AMEM) estimate $\hat{\mu}_{m,n}$ of the measure μ_X to be reconstructed. This estimate is expressed in the form of a discrete measure concentrated on n points of $\mathscr{X}$. In our main result, we prove that $\hat{\mu}_{m,n}$ converges to the solution of the initial inverse problem as $m \to \infty$ and $n \to \infty$ and provide a rate of convergence for this estimate.

The paper is organized as follows. Section 2 introduces some notation and the definition of the AMEM estimate. In Section 3, we state our main result (Theorem 2). Section 4 is devoted to the proofs of our results.

2 Notation and definitions

2.1 Problem position

Let Φ be a continuous and bounded map defined on a subset $\mathscr{X}$ of $\mathbb{R}^d$ and taking values in $\mathbb{R}^k$. The set of finite measures on $(\mathscr{X}, \mathscr{B}(\mathscr{X}))$ will be denoted by $\mathscr{M}(\mathscr{X})$, where $\mathscr{B}(\mathscr{X})$ denotes the Borel σ-field of $\mathscr{X}$. Let $\mu_X \in \mathscr{M}(\mathscr{X})$ be an unknown finite measure on $\mathscr{X}$ and consider the following equation:

$$y = \int_{\mathscr{X}} \Phi(x) d\mu_X(x). \tag{2}$$

Suppose that we observe a perturbed version y^{obs} of the response y:

$$y^{obs} = \int_{\mathscr{X}} \Phi(x) d\mu_X(x) + \varepsilon,$$

where ε is an error term supposed bounded in norm from above by some positive constant η, representing the maximal noise level. Based on the data y^{obs}, we aim at reconstructing the measure μ_X with a maximum entropy procedure. As explained in the introduction, the true map Φ is *unknown* and we assume knowledge of an *approximating sequence* Φ_m to the map Φ, such that

$$\|\Phi_m - \Phi\|_{\mathbb{L}^2(P_X)} = \sqrt{\mathbb{E}(\|\Phi_m(X) - \Phi(X)\|^2)} \to 0,$$

at a rate φ_m.

Let us first introduce some notation. For all probability measure ν on $\mathbb{R}^n$, we shall denote by $\mathscr{L}_\nu$, Λ_ν, and Λ_ν^* the Laplace, log-Laplace, and Cramer transforms of ν, respectively defined for all $s \in \mathbb{R}^n$ by:

$$\mathscr{L}_\nu(s) = \int_{\mathbb{R}^n} \exp\langle s, x\rangle d\nu(x),$$
$$\Lambda_\nu(s) = \log \mathscr{L}_\nu(s),$$
$$\Lambda_\nu^*(s) = \sup_{u \in \mathbb{R}^n} \{\langle s, u\rangle - \Lambda_\nu(u)\}.$$

Define the set
$$K_Y = \{y \in \mathbb{R}^k : \|y - y^{obs}\| \leqslant \eta\},$$
i.e., K_Y is the closed ball centered at the observation y^{obs} and of radius η.

Let $\mathscr{X}$ be a set, and let $\mathscr{P}(\mathscr{X})$ be the set of probability measures on $\mathscr{X}$. For $\nu, \mu \in \mathscr{P}(\mathscr{X})$, the relative entropy of ν with respect to μ is defined by

$$H(\nu|\mu) = \begin{cases} \int_{\mathscr{X}} \log\left(\frac{d\nu}{d\mu}\right) d\nu & \text{if } \nu << \mu \\ +\infty & \text{otherwise.} \end{cases}$$

Given a set $\mathscr{C} \in \mathscr{P}(\mathscr{X})$ and a probability measure $\mu \in \mathscr{P}(\mathscr{X})$, an element $\mu^{\star}$ of $\mathscr{C}$ is called an *I-projection* of μ on $\mathscr{C}$ if

$$H(\mu^{\star}|\mu) = \inf_{\nu \in \mathscr{C}} H(\nu|\mu).$$

Now we let $\mathscr{X}$ be a locally convex topological vector space of finite dimension. The dual of $\mathscr{X}$ will be denoted by $\mathscr{X}'$. The following two Theorems, due to Csiszar (1984), characterize the entropic projection of a given probability measure on a convex set. For their proofs, see Theorem 3 and Lemma 3.3 in Csiszar (1984), respectively.

Theorem 1. *Let μ be a probability measure on $\mathscr{X}$. Let $\mathscr{C}$ be a convex subset of $\mathscr{X}$ whose interior has a non-empty intersection with the convex hull of the support of μ. Let*

$$\Pi(\mathscr{X}) = \{P \in \mathscr{P}(\mathscr{X}) : \int_{\mathscr{X}} x dP(x) \in \mathscr{C}\}.$$

Then the I-projection $\mu^{\star}$ of μ on $\Pi(\mathscr{C})$ is given by the relation

$$d\mu^{\star}(x) = \frac{\exp \lambda^{\star}(x)}{\int_{\mathscr{X}} \exp \lambda^{\star}(u) d\mu(u)} d\mu(x),$$

where $\lambda^{\star} \in \mathscr{X}'$ is given by

$$\lambda^{\star} = \arg \max_{\lambda \in \mathscr{X}'} \left[\inf_{x \in \mathscr{C}} \lambda(x) - \log \int_{\mathscr{X}} \exp \lambda(x) d\mu(x) \right].$$

Now let ν_Z be a probability measure on $\mathbb{R}_+$. Let P_X be a probability measure on $\mathscr{X}$ having full support, and define the convex functional $I_{\nu_Z}(\mu|P_X)$ by:

$$I_{\nu_Z}(\mu|P_X) = \begin{cases} \int_{\mathscr{X}} \Lambda^*_{\nu_Z}\left(\frac{d\mu}{dP_X}\right) dP_X & \text{if } \mu << P_X \\ +\infty & \text{otherwise.} \end{cases}$$

Within this framework, we consider as a solution of the inverse problem (2) a minimizer of the functional $I_{\nu_Z}(\mu|P_X)$ subject to the constraint

$$\mu \in S(K_Y) = \{\mu \in \mathscr{M}(\mathscr{X}) : \int_{\mathscr{X}} \Phi(x) d\mu(x) \in K_Y\}.$$

2.1.1 The AMEM estimate

We introduce the approximate maximum entropy on the mean (AMEM) estimate as a sequence $\hat{\mu}_{m,n}$ of discrete measures on $\mathscr{X}$. In all of the following, the integer m indexes the approximating sequence Φ_m to Φ, while the integer n indexes a random discretization of the space $\mathscr{X}$. For the construction of the AMEM estimate, we proceed as follows.

Let $(X_1, \ldots, X_n)$ be an i.i.d sample drawn from P_X. Thus the empirical measure $\frac{1}{n}\sum_{i=1}^n \delta_{X_i}$ converges weakly to P_X.

Let L_n be the discrete measure with random weights defined by

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

where $(Z_i)_i$ is a sequence of i.i.d. random variables on $\mathbb{R}$.

For $\mathscr{S}$ a set we denote by $\operatorname{co}\mathscr{S}$ its convex hull. Let $\Omega_{m,n}$ be the probability event defined by

$$\Omega_{m,n} = [K_Y \cap \operatorname{co} Supp\, F_* v_Z^{\otimes n} \neq \emptyset] \tag{3}$$

where $F : \mathbb{R}^n \to \mathbb{R}^k$ is the linear operator associated with the matrix $\mathbf{A}_{m,n} = \frac{1}{n}(\Phi_m^i(X_j))_{(i,j)\in[1,k]\times[1,n]}$ and where $F_* v_Z^{\otimes n}$ denotes the image measure of $v_Z^{\otimes n}$ by F. For ease of notation, the dependence of F on m and n will not be explicitly written throughout.

Denote by $\mathscr{P}(\mathbb{R}^n)$ the set of probability measures on $\mathbb{R}^n$. For any map $\Psi : \mathscr{X} \to \mathbb{R}^k$ define the set

$$\Pi_n(\Psi, K_Y) = \left\{ v \in \mathscr{P}(\mathbb{R}^n) : \mathbb{E}_v\left[\int_{\mathscr{X}} \Psi(x)dL_n(x) \right] \in K_Y \right\}.$$

Let $v_{m,n}^\star$ be the I-projection of $v_Z^{\otimes n}$ on $\Pi_n(\Phi_m, K_Y)$.

Then, on the event $\Omega_{m,n}$, we define the AMEM estimate $\hat{\mu}_{m,n}$ by

$$\hat{\mu}_{m,n} = \mathbb{E}_{v_{m,n}^\star}[L_n], \tag{4}$$

and we extend the definition of $\hat{\mu}_{m,n}$ to the whole probability space by setting it to the null measure on the complement $\Omega_{m,n}^c$ of $\Omega_{m,n}$. In other words, letting $(z_1, \ldots, z_n)$ be the expectation of the measure $v_{m,n}^\star$, the AMEM estimate may be rewritten more conveniently as

$$\hat{\mu}_{m,n} = \frac{1}{n}\sum_{i=1}^n z_i \delta_{X_i} \tag{5}$$

with $z_i = \mathbb{E}_{v_{m,n}^\star}(Z_i)$ on $\Omega_{m,n}$, and as $\hat{\mu}_{m,n} \equiv 0$ on $\Omega_{m,n}^c$. It is shown in Loubes and Pelletier (2008) that $\mathbb{P}(\Omega_{m,n}) \to 1$ as $m \to \infty$ and $n \to \infty$. Hence for m and n large enough, the AMEM estimate $\hat{\mu}_{m,n}$ may be expressed as in (5) with high probability, and asymptotically with probability 1.

Remark 1. The construction of the AMEM estimate relies on a discretization of the space $\mathscr{X}$ according to the probability P_X. Therefore by varying the support of P_X, the practitioner may easily incorporate some a-priori knowledge concerning the support of the solution. Similarly, the AMEM estimate also depends on the measure ν_Z, which determines the domain of $\Lambda_{\nu_Z}^*$, and so the range of the solution.

3 Convergence of the AMEM estimate

3.1 Main Result

Assumption 1 The minimization problem admits at least one solution, i.e., there exists a continuous function $g_0 : \mathscr{X} \to \text{co} \, \text{Supp} \, \nu_Z$ such that

$$\int_{\mathscr{X}} \Phi(x) g_0(x) dP_X(x) \in K_Y.$$

Assumption 2

(i) $\text{dom} \, \Lambda_{\nu_Z} := \{s : |\Lambda_{\nu_Z}(s)| < \infty\} = \mathbb{R}$;
(ii) Λ'_{ν_Z} and Λ''_{ν_Z} are bounded.

Assumption 3 The approximating sequence Φ_m converges to Φ in $L^2(\mathscr{X}, P_X)$. Its rate of convergence is given by

$$\|\Phi_m - \Phi\|_{\mathbb{L}^2} = O(\varphi_m^{-1})$$

Assumption 4 Λ_{ν_Z} is a convex function

Assumption 5 For all m, the components of Φ_m are linearly independent

Assumption 6 Λ'_{ν_Z} and Λ''_{ν_Z} are continuous functions.
We are now in a position to state our main result.

Theorem 2 (Convergence of the AMEM estimate). *Suppose that Assumption 1, Assumption 2, and Assumption 3 hold. Let μ^* be the minimizer of the functional*

$$I_{\nu_Z}(\mu|P_X) = \int_{\mathscr{X}} \Lambda_{\nu_Z}^* \left(\frac{d\mu}{dP_X} \right) dP_X$$

subject to the constraint $\mu \in S(K_Y) = \{\mu \in \mathscr{M}(\mathscr{X}) : \int_{\mathscr{X}} \Phi(x) d\mu(x) \in K_Y\}$.

- *Then the AMEM estimate $\hat{\mu}_{m,n}$ is defined by*

$$\hat{\mu}_{m,n} = \frac{1}{n}\sum_{i=1}^{n}\Lambda'_{v_Z}(\langle \hat{v}_{m,n}, \Phi_m(X_i)\rangle)\delta_{X_i}$$

where $\hat{v}_{m,n}$ minimizes on $\mathbb{R}^k$

$$H_n(\Phi_m,v) = \frac{1}{n}\sum_{i=1}^{n}\Lambda_{v_Z}(\langle v, \Phi_m(X_i)\rangle) - \inf_{y\in K_Y}\langle v,y\rangle$$

- *Moreover, under Assumption 4, Assumption 2, and Assumption 3, it converges weakly to μ^* as $m\to\infty$ and $n\to\infty$. Its rate of convergence is given by*

$$\|\hat{\mu}_{m,n} - \mu^*\|_{VT} = O_P(\varphi_m^{-1}) + O_P\left(\frac{1}{\sqrt{n}}\right).$$

Remark 2. Assumption 2-(i) ensures that the function $H(\Phi,v)$ in Theorem 2 attains its minimum at a unique point $v^\star$ belonging to the interior of its domain. If this assumption is not met, Borwein and Lewis (1993) and Gamboa and Gassiat (1999) have shown that the minimizers of $I_{v_Z}(\mu|P_X)$ over $S(K_Y)$ may have a singular part with respect to P_X.

Proof. The rate of convergence of the AMEM estimate depends both on the discretization n and the convergence of the approximated operator m. Hence we consider

$$\hat{v}_{m,\infty} = \underset{v\in\mathbb{R}^k}{\operatorname{argmin}} H(\Phi_m,v) = \underset{v\in\mathbb{R}^k}{\operatorname{argmin}}\left\{\int_{\mathscr{X}}\Lambda_{v_Z}(\langle \Phi_m(x),v\rangle)dP_X - \inf_{y\in K_Y}\langle v,y\rangle\right\},$$

$$\hat{\mu}_{m,n} = \frac{1}{n}\sum_{i=1}^{n}\Lambda'_{v_Z}(\langle \hat{v}_{m,n},\Phi_m(.)\rangle)\delta_{X_i},$$

$$\hat{\mu}_{m,\infty} = \Lambda'_{v_Z}(\langle \Phi_m(.),\hat{v}_{m,\infty}\rangle)P_X.$$

We have the following upper bound

$$\|\hat{\mu}_{m,n} - \mu^*\|_{VT} \leqslant \|\hat{\mu}_{m,n} - \hat{\mu}_{m,\infty}\|_{VT} + \|\hat{\mu}_{m,\infty} - \mu^*\|_{VT},$$

where each term must be tackled separately.
First, let us consider $\|\hat{\mu}_{m,n} - \hat{\mu}_{m,\infty}\|_{VT}$.

$$\begin{aligned}\|\hat{\mu}_{m,n} - \hat{\mu}_{m,\infty}\|_{VT} =& \|\frac{1}{n}\sum_{i=1}^{n}\Lambda'_{v_Z}(\langle \Phi_m,\hat{v}_{m,n}\rangle)\delta_{X_i} - \Lambda'_{v_Z}(\langle \Phi_m,\hat{v}_{m,\infty}\rangle)P_X\|_{VT}\\ \leqslant& \|\frac{1}{n}\sum_{i=1}^{n}\left(\Lambda'_{v_Z}(\langle \Phi_m,\hat{v}_{m,n}\rangle) - \Lambda'_{v_Z}(\langle \Phi_m,\hat{v}_{m,\infty}\rangle)\right)\delta_{X_i}\|_{VT}\\ &+ \|\frac{1}{n}\sum_{i=1}^{n}\Lambda'_{v_Z}(\langle \Phi_m,\hat{v}_{m,\infty}\rangle)\delta_{X_i} - \Lambda'_{v_Z}(\langle \Phi_m,\hat{v}_{m,\infty}\rangle)P_X\|_{VT}\end{aligned}$$

To bound the first term $\|\frac{1}{n}\sum_{i=1}^{n}\left(\Lambda'_{V_Z}(\langle\Phi_m,\hat{v}_{m,n}\rangle)-\Lambda'_{V_Z}(\langle\Phi_m,\hat{v}_{m,\infty}\rangle)\right)\delta_{X_i}\|_{VT}$, let g be a bounded measurable function and write

$$\frac{1}{n}\sum_{i=1}^{n}g(X_i)\left(\Lambda'_{V_Z}(\langle\Phi_m(X_i),\hat{v}_{m,n}\rangle)-\Lambda'_{V_Z}(\langle\Phi_m(X_i),\hat{v}_{m,\infty}\rangle)\right)$$

$$\leqslant\|g\|_\infty\|\Lambda''_{V_Z}\|_\infty\frac{1}{n}\sum_{i=1}^{n}\langle\Phi_m(X_i),\hat{v}_{m,n}-\hat{v}_{m,\infty}\rangle$$

$$\leqslant\|g\|_\infty\|\Lambda''_{V_Z}\|_\infty\|\hat{v}_{m,n}-\hat{v}_{m,\infty}\|\frac{1}{n}\sum_{i=1}^{n}\|\Phi_m(X_i)\|$$

where we have used Cauchy-Schwarz inequality. Since $(\Phi_m)_m$ converges in $\mathbb{L}^2(P_X)$, it is bounded in $\mathbb{L}^2(P_X)$-norm, yelding that $\frac{1}{n}\sum_{i=1}^{n}\|\Phi_m(X_i)\|$ converges almost surely to $\mathbb{E}\|\Phi_m(X)\|<\infty$. Hence, there exists $K_1>0$ such that

$$\|\frac{1}{n}\sum_{i=1}^{n}\left(\Lambda'_{V_Z}(\langle\Phi_m,\hat{v}_{m,n}\rangle)-\Lambda'_{V_Z}(\langle\Phi_m,\hat{v}_{m,\infty}\rangle)\right)\delta_{X_i}\|_{VT}\leqslant K_1\|\hat{v}_{m,n}-\hat{v}_{m,\infty}\|.$$

For the second term, we obtain

$$\|\frac{1}{n}\sum_{i=1}^{n}\Lambda'_{V_Z}(\langle\Phi_m,\hat{v}_{m,\infty}\rangle)\delta_{X_i}-\Lambda'_{V_Z}(\langle\Phi_m,\hat{v}_{m,\infty}\rangle)P_X\|_{VT}=O_P\left(\frac{1}{\sqrt{n}}\right).$$

Hence we get

$$\|\hat{\mu}_{m,n}-\hat{\mu}_{m,\infty}\|_{VT}\leqslant K_1\|\hat{v}_{m,n}-\hat{v}_{m,\infty}\|+O_P\left(\frac{1}{\sqrt{n}}\right).$$

The second step is to consider $\|\hat{\mu}_{m,\infty}-\mu^*\|_{VT}$ and to follow the same guidelines. So, we get

$$\|\hat{\mu}_{m,\infty}-\mu^*\|_{VT}=\|\left(\Lambda'_{V_Z}(\langle\Phi_m,\hat{v}_{m,\infty}\rangle)-\Lambda'_{V_Z}(\langle\Phi,v^*\rangle)\right)P_X\|_{VT}$$

$$\leqslant\|\left(\Lambda'_{V_Z}(\langle\Phi_m,\hat{v}_{m,\infty}\rangle)-\Lambda'_{V_Z}(\langle\Phi_m,v^*\rangle)\right)P_X\|_{VT}$$

$$+\|\left(\Lambda'_{V_Z}(\langle\Phi_m,v^*\rangle)-\Lambda'_{V_Z}(\langle\Phi,v^*\rangle)\right)P_X\|_{VT}$$

Fo any bounded measurable function g, we can write still using Cauchy-Schwarz inequality that

$$\int_{\mathscr{X}}g(x)\left(\Lambda'_{V_Z}(\langle\Phi_m(x),\hat{v}_{m,\infty}\rangle)-\Lambda'_{V_Z}(\langle\Phi_m(x),v^*\rangle)\right)dP_X(x)$$

$$\leqslant\int_{\mathscr{X}}g(x)\Lambda''_{V_Z}(\xi)\langle\Phi_m(x),\hat{v}_{m,\infty}-v^*\rangle dP_X(x)$$

$$\leqslant\|\Lambda''_{V_Z}\|_\infty\sqrt{\mathbb{E}(g(X))^2}\sqrt{\mathbb{E}(\|\Phi_m(X)\|^2)}\,\|\hat{v}_{m,\infty}-v^*\|$$

Hence there exists $K_2 > 0$ such that

$$\| \left(\Lambda'_{v_Z}(\langle \Phi_m, \hat{v}_{m,\infty} \rangle) - \Lambda'_{v_Z}(\langle \Phi_m, v^* \rangle) \right) P_X \|_{VT} \leqslant K_2 \| \hat{v}_{m,\infty} - v^* \|.$$

Finally, the last term $\| \left(\Lambda'_{v_Z}(\langle \Phi_m, v^* \rangle) - \Lambda'_{v_Z}(\langle \Phi, v^* \rangle) \right) P_X \|_{VT}$ can be bounded. Indeed, for any measurable bounded g

$$\int_{\mathcal{X}} g(x) \left(\Lambda'_{v_Z}(\langle \Phi_m(x), v^* \rangle) - \Lambda'_{v_Z}(\langle \Phi(x), v^* \rangle) \right) dP_X(x)$$
$$= \int_{\mathcal{X}} g(x) \Lambda''_{v_Z}(\xi_x) \langle \Phi_m(x) - \Phi(x), v^* \rangle dP_X(x)$$
$$\leqslant \int_{\mathcal{X}} g(x) \Lambda''_{v_Z}(\xi_x) \| \Phi_m(x) - \Phi(x) \| \| v^* \| dP_X(x)$$
$$\leqslant \| v^* \| \| \Lambda''_{v_Z} \|_\infty \sqrt{\mathbb{E}(g(X))^2} \sqrt{\mathbb{E}(\| \Phi_m(X) - \Phi(X) \|^2)}$$

Hence there exists $K_3 > 0$ such that

$$\| \left(\Lambda'_{v_Z}(\langle \Phi_m, v^* \rangle) - \Lambda'_{v_Z}(\langle \Phi, v^* \rangle) \right) P_X \|_{VT} \leqslant K_3 \| \Phi_m - \Phi \|_{\mathbb{L}^2}$$

We finally obtain the following bound

$$\| \hat{\mu}_{m,n} - \mu^* \|_{VT} \leqslant K_1 \| \hat{v}_{m,n} - \hat{v}_{m,\infty} \| + K_2 \| \hat{v}_{m,\infty} - v^* \| + K_3 \| \Phi_m - \Phi \|_{\mathbb{L}^2} + O_P \left(\frac{1}{\sqrt{n}} \right)$$

Using Lemmas 1 and 2, we obtain that

$$\| \hat{v}_{m,n} - \hat{v}_{m,\infty} \| = O_P \left(\frac{1}{\sqrt{n}} \right)$$

$$\| \hat{v}_{m,\infty} - v^* \| = O_P(\varphi_m^{-1})$$

Finally, we get

$$\| \hat{\mu}_{m,n} - \mu^* \|_{VT} = O_P(\varphi_m^{-1}) + O_P \left(\frac{1}{\sqrt{n}} \right),$$

which proves the result. $\square$

3.2 Application to remote sensing

In remote sensing of aerosol vertical profiles, one wishes to recover the concentration of aerosol particules from noisy observations of the radiance field (i.e., a radiometric quantity), in several spectral bands (see e.g. Gabella et al, 1997; Gabella, Kisselev and Perona, 1999). More specifically, at a given level of modeling, the noisy observation y^{obs} may be expressed as

$$y^{obs} = \int_{\mathscr{X}} \Phi(x;t^{obs})d\mu_X(x) + \varepsilon, \tag{6}$$

where $\Phi : \mathscr{X} \times \mathscr{T} \to \mathbb{R}^k$ is a given operator, and where t^{obs} is a vector of angular parameters observed simultaneously with y^{obs}. The aerosol vertical profile is a function of the altitude x and is associated with the measure μ_X to be recovered, i.e., the aerosol vertical profile is the Radon-Nykodim derivative of μ_X with respect to a given reference measure (e.g., the Lebesgue measure on $\mathbb{R}$). The analytical expression of Φ is fairly complex as it sums up several models at the microphysical scale, so that basically Φ is available in the form of a computer code. So this problem motivates the introduction of an efficient numerical procedure for recovering the unknwon μ_X from y^{obs} and arbitrary t^{obs}.

More generally, the remote sensing of the aerosol vertical profile is in the form of an inverse problem where some of the inputs (namely t^{obs}) are observed simultaneously with the noisy output y^{obs}. Suppose that random points $X_1, \ldots, X_n$ of $\mathscr{X}$ have been generated. Then, applying the maximum entropy approach would require the evaluations of $\Phi(X_i, t^{obs})$ each time t^{obs} is observed. If one wishes to process a large number of observations, say (y_i^{obs}, t_i^{obs}), for different values t_i^{obs}, the computational cost may become prohibitive. So we propose to replace Φ by an approximation Φ_m, the evaluation of which is faster in execution. To this aim, suppose first that $\mathscr{T}$ is a subset of $\mathbb{R}^p$. Let $T_1, \ldots, T_m$ be random points of $\mathscr{T}$, independent of $X_1, \ldots, X_n$, and drawn from some probability measure μ_T on $\mathscr{T}$ admitting a density f_T with respect to the Lebesgue measure on $\mathbb{R}^p$ such that $f_T(t) > 0$ for all $t \in \mathscr{T}$. Next, consider the operator

$$\Phi_m(x,t) = \frac{1}{f_T(t)} \frac{1}{m} \sum_{i=1}^{m} K_{h_m}(t - T_i)\Phi(x, T_i),$$

where $K_{h_m}(.)$ is a symetric kernel on $\mathscr{T}$ of smoothing sequence h_n. It is a classical result to prove that Φ_m converges to Φ in quadratic norm provided h_m tends to 0 at a suitable rate, which ensures that Assumption 3 of Theorem 2 is satisfied. Since the T_i's are independent from the X_i, one may see that Theorem 2 applies, and so the solution to the approximate inverse problem

$$y^{obs} = \int_{\mathscr{X}} \Phi_m(x;t^{obs})d\mu_X(x) + \varepsilon,$$

will converge to the solution to the original inverse problem in Eq. 6. In terms of computational complexity, the advantage of this approach is that the construction of the AMEM estimate requires, for each new observation (y^{obs}, t^{obs}), the evaluation of the m kernels at t^{obs}, i.e., $K_{h_m}(t^{obs} - T_i)$, the $m \times n$ ouputs $\Phi(X_i, T_j)$ for $i = 1, \ldots, n$ and $j = 1, \ldots, m$ having evaluated once and for all.

3.3 Application to deconvolution type problems in optical nanoscopy

Following the framework defined in [17], the number of photons counted can be expressed using a convolution of $p(x-y,y)$ the probability of recording a photon emission at point y when illuminating point x, with $d\mu(y) = f(y)dy$ the measure of the fluorescent markers.

$$g(x) = \int p(x-y,x)f(y)dy.$$

Here $p(x-y,y) = p(x,y,\phi(x))$. Reconstruction of μ can be achieved using AMEM technics.

4 Tecnical Lemmas

Recall the following definitions

$$\hat{v}_{m,\infty} = \operatorname*{argmin}_{v\in\mathbb{R}^k} H(\Phi_m,v) = \operatorname*{argmin}_{v\in\mathbb{R}^k}\left\{\int_{\mathscr{X}}\Lambda_{V_Z}(\langle\Phi_m(x),v\rangle)dP_X - \inf_{y\in K_Y}\langle v,y\rangle\right\},$$

$$\hat{v}_{m,n} = \operatorname*{argmin}_{v\in\mathbb{R}^k} H_n(\Phi_m,v) = \operatorname*{argmin}_{v\in\mathbb{R}^k}\left\{\frac{1}{n}\sum_{i=1}^{n}\Lambda_{V_Z}(\langle v,\Phi_m(X_i)\rangle) - \inf_{y\in K_Y}\langle v,y\rangle\right\},$$

$$v^* = \operatorname*{argmin}_{v\in\mathbb{R}^k} H(\Phi,v) = \operatorname*{argmin}_{v\in\mathbb{R}^k}\left\{\int_{\mathscr{X}}\Lambda_{V_Z}(\langle\Phi(x),v\rangle)dP_X(x) - \inf_{y\in K_Y}\langle v,y\rangle\right\}$$

Lemma 1 (Uniform convergence at a given approximation level m). *For all m, we get*

$$\|\hat{v}_{m,n} - \hat{v}_{m,\infty}\| = O_P\left(\frac{1}{\sqrt{n}}\right)$$

Proof. $\hat{v}_{m,n}$ is defined as the minimizer of an empirical constrast function $H_n(\Phi_m,.)$. Indeed, set

$$h_m(v,x) = \Lambda_{V_Z}(\langle\Phi_m(x),v\rangle) - \inf_{y\in K_Y}\langle v,y\rangle,$$

hence

$$H(\Phi_m,v) = P_X h_m(v,.).$$

Using classical theorem from the theory of M-estimation, we get the convergence in probability of $\hat{v}_{m,n}$ towards $\hat{v}_{m,\infty}$ provided that the contrast converges uniformly over every compact set of $\mathbb{R}^k$ towards $H(\Phi_m,.)$ when $n \to \infty$. More precisely Corollary 5.53 in van der Vaart (1998) states that if we consider $x \mapsto h_m(v,x)$ a measurable function and $\dot{h}_m$ a function in $L^2(P)$, such that for all v_1 and v_2 in a neighbourhood of v^*

$$|h_m(v_1, x) - h_m(v_2, x)| \leqslant \dot{h}_m(x)\|v_1 - v_2\|.$$

Moreover if $v \mapsto Ph_m(v, .)$ has a Taylor expansion of order at least 2 around its unique minimum v^* and if the Hessian matrix at this point is positive, hence provided $\mathbb{P}_n h_m(\hat{v}_n, .) \leqslant \mathbb{P}_n h_m(v^*,) + O_P(n^{-1})$ then

$$\sqrt{n}(\hat{v}_n - v^*) = O_P(1).$$

We want to apply this result to our problem. Let η be an un upper bound for $\|\varepsilon\|$, we set $h_m(v, x) = \Lambda_{v_Z}(\langle \Phi_m(x), v \rangle) - \langle v, y^{obs} \rangle - \inf_{\|y - y^{obs}\| \leqslant \eta} \langle v, y - y_{obs} \rangle$. Now note that $z \mapsto \langle v, z \rangle$ reaches its minimum on $\mathscr{B}(0, \eta)$ at the point $-\eta \dfrac{v}{\|v\|}$, so

$$h_m(v, x) = \Lambda_{v_Z}(\langle \Phi_m(x), v \rangle) - \langle v, y^{obs} \rangle + \eta\|v\|$$

For all $v_1, v_2 \in \mathbb{R}^k$, we have

$$\begin{aligned}
&|h_m(v_1, x) - h_m(v_2, x)| \\
&= |\Lambda_{v_Z}(\langle \Phi_m(x), v_1 \rangle) - \inf_{y \in K_Y} \langle v_1, y \rangle - \Lambda_{v_Z}(\langle \Phi_m(x), v_2 \rangle) + \inf_{y \in K_Y} \langle v_2, y \rangle| \\
&\leqslant |\Lambda_{v_Z}(\langle \Phi_m(x), v_1 \rangle) - \Lambda_{v_Z}(\langle \Phi_m(x), v_2 \rangle)| + |\inf_{y \in K_Y} \langle v_2, y \rangle - \inf_{y \in K_Y} \langle v_1, y \rangle| \\
&\leqslant |\Lambda_{v_Z}(\langle \Phi_m(x), v_1 \rangle) - \Lambda_{v_Z}(\langle \Phi_m(x), v_2 \rangle)| + |\langle v_2 - v_1, y^{obs} \rangle - \eta(\|v_2\| - \|v_1\|)| \\
&\leqslant \left(\|\Lambda'_{v_Z}\|_\infty \|\Phi_m(x)\| + \|y^{obs}\| + \eta \right) \|v_1 - v_2\|
\end{aligned}$$

Define $\dot{h}_m : x \mapsto \|\Lambda'_{v_Z}\|_\infty \|\Phi_m(x)\| + \|y^{obs}\| + \eta$. Since $(\Phi_m)_m$ is bounded in $\mathbb{L}^2(P_X)$, $(\dot{h}_m)_m$ is in $L^2(P_X)$ uniformly with respect to m, which entails that

$$\exists K, \forall m, \int_{\mathscr{X}} \dot{h}_m^2 dP_X < K \tag{7}$$

Hence the function $\dot{h}_m$ satisifes the first condition

$$|h_m(v_1, x) - h_m(v_2, x)| \leqslant \dot{h}_m(x)\|v_1 - v_2\|$$

Now, consider $H(\Phi_m, .)$ Let $V_{m,v}$ be the Hessian matrix of $H(\Phi_m, .)$ at point v. We need to prove that $V_{m, \hat{v}_{m,\infty}}$ is non negative. Let ∂_i be the derivative with respect to the ith component. Set $v \neq 0$, we have

$$\begin{aligned}
V_{m,v}^{ij}(v) = \partial_i \partial_j H(\Phi_m, v) &= \int_{\mathscr{X}} \partial_i \partial_j h_m(v, x) dP_X \\
&= \int_{\mathscr{X}} \Phi_m^i(x) \Phi_m^j(x) \Lambda''_{v_Z}(\langle \Phi_m(x), v \rangle) dP_X + \eta \, \partial_i \partial_j N(v)
\end{aligned}$$

where let N be $N : v \mapsto \|v\|$.

Hence the Hessian matrix $V_{m, \hat{v}_{m,\infty}}$ of $H(\Phi_m, .)$ at point $\hat{v}_{m,\infty}$ can be split into the sum

ot the following matrices

$$(M_1)_{ij} = \int_{\mathscr{X}} \Phi_m^i(x)\Phi_m^j(x)\Lambda_{\nu_Z}''(\langle\Phi_m(x),\hat{v}_{m,\infty}\rangle)dP_X,$$
$$(M_2)_{ij} = \partial_i\partial_j N(\hat{v}_{m,\infty}).$$

Under Assumptions (A3) and (A5), Λ_{ν_Z}'' is positive and belongs to $L_1(P_X)$ since it is bounded. So we can define $\int_{\mathscr{X}} \Phi_m^i(x)\Phi_m^j(x)\Lambda_{\nu_Z}''(\langle\Phi_m(x),\hat{v}_{m,\infty}\rangle)dP_X$ as the scalar product of Φ_m^i and Φ_m^j in the space $\mathbb{L}^2(\Lambda_{\nu_Z}''(\langle\Phi_m(.),\hat{v}_{m,\infty}\rangle P_X)$.
M_1 is a Gram matrix, hence using (A6) it is a non negative matrix.
M_2 can be computed as follows. For all $v \in \mathbb{R}^k/\{0\}$, we have

$$N(v) = \sqrt{\Sigma_{i=1}^k v_i^2}$$
$$\partial_i N(v) = \frac{v_i}{\|v\|}$$
$$\partial_i\partial_j N(v) = \begin{cases} -\dfrac{v_i v_j}{\|v\|^3} & \text{si } i \neq j \\ \dfrac{\|v\|^2 - v_i^2}{\|v\|^3} & \text{si } i = j \end{cases}$$

Hence for all $a \in \mathbb{R}^k$, we can write

$$\begin{aligned}
&a^T M_2 a \\
&= \sum_{1 \leqslant i,j \leqslant k} \partial_i\partial_j N(\hat{v}_{m,\infty})a_i a_j \\
&= \sum_{i=1}^k \frac{\|\hat{v}_{m,\infty}\|^2 - \hat{v}_{m,\infty,i}^2}{\|\hat{v}_{m,\infty}\|^3}a_i^2 - \sum_{i \neq j} \frac{\hat{v}_{m,\infty,i}\hat{v}_{m,\infty,j}}{\|\hat{v}_{m,\infty}\|^3}a_i a_j \\
&= \frac{1}{\|\hat{v}_{m,\infty}\|^3}\left(\|\hat{v}_{m,\infty}\|^2\sum_{i=1}^k a_i^2 - \sum_{i=1}^k a_i^2\hat{v}_{m,\infty,i}^2 - \sum_{1 \leqslant i,j \leqslant k} a_i\hat{v}_{m,\infty,i}a_j\hat{v}_{m,\infty,j} + \sum_{i=1}^k a_i^2\hat{v}_{m,\infty,i}^2\right) \\
&= \frac{1}{\|\hat{v}_{m,\infty}\|^3}\left(\|\hat{v}_{m,\infty}\|^2\|a\|^2 - \sum_{1 \leqslant i,j \leqslant k} a_i\hat{v}_{m,\infty,i}a_j\hat{v}_{m,\infty,j}\right) \\
&= \frac{1}{\|\hat{v}_{m,\infty}\|^3}\left(\|\hat{v}_{m,\infty}\|^2\|a\|^2 - \langle a,\hat{v}_{m,\infty}\rangle^2\right) \geqslant 0 \text{ using Cauchy-Schwarz's inequality.}
\end{aligned}$$

So M_2 is clearly non negative, hence $V_{m,\hat{v}_{m,\infty}} = M_1 + \eta M_2$ is also non negative. Finally we conclude that $H(\Phi_m,.)$ undergoes the assumptions of Theorem 5.1. $\square$

Lemma 2.
$$\|\hat{v}_{m,\infty} - v^*\| = O_P(\varphi_m^{-1})$$

Proof. First write,

$$|H(\Phi_m,v) - H(\Phi,v)| = |\int_{\mathscr{X}} \Lambda_{vZ}(\langle \Phi_m(x),v\rangle) - \Lambda_{vZ}(\langle \Phi(x),v\rangle)dP_X(x)|$$
$$\leqslant \|\Lambda'_{vZ}\|_\infty \|v\| \|\Phi_m - \Phi\|_{\mathbb{L}^2},$$

which implies uniform convergence over every compact set of $H(\Phi_m,.)$ towards $H(\Phi,.)$ when $m \to \infty$, yelding that $\hat{v}_{m,\infty} \to v^*$ in probability. To compute the rate of convergence, we use Lemma 3. As previously we can show that the Hessian matrix of $H(\phi,.)$ at point v^* is positive. We need to prove uniform convergence of $\nabla H(\phi_m,.)$ towards $\nabla H(\phi,.)$. For this, write

$$\partial_i\left[H(\phi_m,.) - H(\phi,.)\right](v)$$
$$= \int_{\mathscr{X}} \Phi_m^i(x)\Lambda'_{vZ}(\langle \Phi_m(x),v\rangle) - \Phi^i(x)\Lambda'_{vZ}(\langle \Phi(x),v\rangle)dP_X(x)$$
$$= \int_{\mathscr{X}} (\Phi_m^i - \Phi^i)(x)\Lambda'_{vZ}(\langle \Phi_m(x),v\rangle) - \Phi^i(x)\Lambda''_{vZ}(\xi)\langle (\Phi - \Phi_m)(x),v\rangle dP_X(x)$$
$$\leqslant \|\Phi^i - \Phi_m^i\|_{\mathbb{L}^2}\|\Lambda'_{vZ}\|_\infty + \|\Phi^i\|_{\mathbb{L}^2}\|\Lambda''_{vZ}\|_\infty\|\Phi - \Phi_m\|_{\mathbb{L}^2}\|v\|$$

using again Cauchy-Schwarz's inequality. Finally we obtain

$$\|\nabla(H(\phi_m,.) - H(\Phi,.))(v)\| \leqslant (C_1 + C_2\|v\|)\|\Phi - \Phi_m\|_{\mathbb{L}^2}$$

for positive constants C_1 and C_2. For any compact neighbourhood of v^*, $\mathscr{S}$, the function $v \mapsto \|\nabla(H(\phi_m,.) - H(\Phi,.))(v)\|$ converges uniformly to 0. But for m large enough, $\hat{v}_{m,\infty} \in \mathscr{S}$ almost surely. Using 2. in Lemma 3 with the function $v \mapsto \|\nabla(H(\phi_m,.) - H(\Phi,.))(v)\|\mathbf{1}_{\mathscr{S}}(v)$ converging uniformly to 0, implies that

$$\|\hat{v}_{m,\infty} - v^*\| = O_P(\varphi_m^{-1}). \square$$

Lemma 3. *Let f be defined on $\mathscr{S} \subset \mathbb{R}^d \to \mathbb{R}$, which reaches a unique minimum at point θ_0. Let $(f_n)_n$ be a sequence of continuous functions which converges uniformly towards f. Let $\hat{\theta}_n = \arg\min f_n$. If f is twice differentiable on a neighbourhood of θ_0 and provided its Hessian matrix V_{θ_0} is non negative, hence we get*

1. there exists a positive constant C such that

$$\|\hat{\theta}_n - \theta_0\| \leqslant C\sqrt{\|f - f_n\|_\infty}$$

2. Moreover if $\theta \mapsto V_\theta$ is continuous in a neighbourhood of θ_0 and $\|\nabla f_n(.)\|$ uniformly converges towards $\|\nabla f(.)\|$, hence there exists a constant C' such that

$$\|\hat{\theta}_n - \theta_0\| \leqslant C'\|\nabla(f - f_n)\|_\infty$$

with $\|g\|_\infty = \sup_{x \in \mathscr{S}} \|g(x)\|$

Proof. The proof of this classical result in optimization relies on easy convex analysis tricks. For sake of completeness, we recall here the main guidelines.
1. There are non negative constants C_1 et δ_0 such that

$$\forall\, 0 < \delta \leqslant \delta_0, \quad \inf_{d(\theta,\theta_0)>\delta} f(\theta) - f(\theta_0) > C_1 \delta^2$$

Set $\|f_n - f\|_\infty = \varepsilon_n$. For $0 < \delta_1 < \delta_0$, let n be chosen such that $2\varepsilon_n \leqslant C_1 \delta_1^2$. Hence

$$\inf_{d(\theta,\theta_0)>\delta_1} f_n(\theta) \geqslant \inf_{d(\theta,\theta_0)>\delta_1} f(\theta) - \varepsilon_n > f(\theta_0) + \varepsilon_n \geqslant f_n(\theta_0)$$

Finally $f_n(\theta_0) < \inf\limits_{d(\theta,\theta_0)>\delta_1} f_n(\theta) \implies \hat{\theta}_n \in \{\theta : d(\theta,\theta_0) \leqslant \delta_1\}$, which enables to conclude setting $C = \sqrt{\frac{2}{C_1}}$.

2. We prove the result for $d = 1$, which can be easily extended for all d. Using Taylor-Lagrange expansion, there exists $\tilde{\theta}_n \in\,]\hat{\theta}_n, \theta_0[$ such that

$$f'(\theta_0) = 0 = f'(\hat{\theta}_n) + (\theta_0 - \hat{\theta}_n) f''(\tilde{\theta}_n).$$

Remind that $f''(\tilde{\theta}_n) \xrightarrow[n\to\infty]{} f''(\theta_0) > 0$. So, for n large enough there exits $C' > 0$ such that

$$|\theta_0 - \hat{\theta}_n| = \frac{|f'(\hat{\theta}_n) - f'(\theta_0)|}{|f''(\tilde{\theta}_n)|} \leqslant C' \|f' - f_n'\|_\infty,$$

which ends the proof. $\quad\square$

References

1. Borwein, J.M., Lewis A.S. Partially-finite programming in l^1 and the existence of maximum entropy estimates. *SIAM J. Optim.*, 3:248–267, 1993.
2. Borwein, J.M., Lewis A.S., Noll, D. Maximum entropy reconstruction using derivative information i: Fisher information and convex duality. *Math. Oper. Res.*, 21:442–468, 1996.
3. Carasco, M., Florens J.P., Renault, E. Linear inverse problems in structural econometrics: Estimation based on spectral decomposition and regularization. In *Handbook of Econometrics*, J.J. Heckman and E.E. Leamer, Eds, North Holand, 2006.
4. Cavalier, L., Hengartner, N.W. Adaptive estimation for inverse problems with noisy operators. *Inverse Problems*, 21:1345–1361, 2005.
5. Csiszár,I. Sanov property, generalized *I*-projection and a conditional limit theorem. *Ann. Probab.*, 12(3):768–793, 1984.
6. Decarreau, A., Hilhorst, D., Lemaréchal, C., Navaza,J. Dual methods in entropy maximization. Application to some problems in crystallography. *SIAM J. Optim.*, 2(2):173–197, 1992.
7. Eggermont, P.P.B. Maximum entropy regularization for Fredholm integral equations of the first kind. *SIAM J. Math. Anal.*, 24(6):1557–1576, 1993.
8. Efromovich, S., Koltchinskii, C. On inverse problems with unknown operators. *IEEE Trans. Inform. Theory*, 47(7):2876–2894, 2001.
9. Engl, H.W., Hanke, M., Neubauer, A. *Regularization of inverse problems*, volume 375 of *Mathematics and its Applications*. Kluwer Academic Publishers Group, Dordrecht, 1996.
10. Evans, S.N., Stark, P.B. Inverse problems as statistics. *Inverse Problems*, 18(4):55-97, 2002.
11. Hall, P., Horowitz,J.L. Nonparametric methods for inference in the presence of instrumental variables. *The Annals of Statistics*, 33(6):2904-2929, 2005.

12. Fermin, A.K., Loubes, J.M., Ludeña, C. Bayesian methods in seismic tomography. *International Journal of Tomography and Statistics*, 4(6):1–19, 2006.

13. Gabella, M., Guzzi, R., Kisselev, V., Perona, G. Retrieval of aerosol profile variations in the visible and near infrared: theory and application of the single-scattering approach. *Applied Optics*, 36(6):1328-1336, 1997.

14. Gabella, M., Kisselev, V., Perona, G. Retrieval of aerosol profile variations from reflected radiation in the oxygen absorption A band. *Applied Optics*, 38(15):3190-3195, 1999.

15. Gamboa, F. New Bayesian methods for ill posed problems. *Stat. Decis.*, 17(4):315–337, 1999.

16. Gamboa, F., Gassiat, E. Bayesian methods and maximum entropy for ill-posed inverse problems. *Ann. Statist.*, 25(1):328–350, 1997.

17. Giewekemeyer, T., Hohage, K., Salditt, T. Iterative reconstruction of a refractive index from x-ray or neutron reflectivity measurements. *Physical Review E.*, 77:051604, 2008.

18. Gzyl, H. Tomographic reconstruction by maximum entropy in the mean: unconstrained reconstructions. *Appl. Math. Comput.*, 129(2-3):157–169, 2002.

19. Gzyl, H., Zeev, N. Probabilistic approach to an image reconstruction problem. *Methodol. Comput. Appl. Probab.*, 4(3):279–290, 2003.

20. Hermann, U., Noll, D. Adaptive image reconstruction using information measures. *SIAM J. Control Optim.*, 38(4):1223–1240, 2000.

21. Rockafellar, R.T. *Convex analysis*. Princeton Landmarks in Mathematics. Princeton University Press, Princeton, NJ, 1997. Reprint of the 1970 original, Princeton Paperbacks.

22. Skilling, J. Maximum entropy spectroscopy—DIMES and MESA. In *Maximum-entropy and Bayesian methods in science and engineering, Vol. 2 (Laramie, WY, 1985 and Seattle, WA, 1986/1987)*, Fund. Theories Phys., pages 127–145. Kluwer Acad. Publ., Dordrecht, 1988.

23. Skilling, J., Gull, S. Bayesian maximum entropy image reconstruction. In *Spatial statistics and imaging (Brunswick, ME, 1988)*, volume 20 of *IMS Lecture Notes Monogr. Ser.*, pages 341–367. Inst. Math. Statist., Hayward, CA, 1991.

24. van der Vaart, A.W. "Asymptotics Statistics", *Cambr. Univ. Press*, 1998.