Statistical minimax approach of the Hausdorff moment problem

Thanh Mai Pham Ngoc

Université Pierre et Marie Curie, Paris VI

Abstract: The purpose of this paper is to study the problem of estimating a compactly supported function from noisy observations of its moments in an ill-posed inverse problem framework. We provide a statistical approach to the famous Hausdorff classical moment problem. We prove a lower bound on the rate of convergence of the mean integrated squared error and provide an estimator which attains minimax rate over the corresponding smoothness classes.

Key words and phrases: Minimax estimation, Statistical inverse problems, Problem of the moments, Legendre polynomials, Kullback-Leibler information

MSC 2000 Subject Classification 62G07 44A60 45Q05

1. Introduction

The classical moment problem can be stated as follows: it consists in getting some information about a distribution μ from the knowledge of its moments $\int x^k d\mu(x)$. This problem has been largely investigated in many mathematical topics, among others, in operator theory, mathematical physics, inverse spectral theory, probability theory, inverse problems, numerical analysis but also in a wide variety of settings in physics and engineering such as quantum spin models, speech processing, radio astronomy (see for instance Kay and Marple (1981), Lang and McClellan (1983) ...). We may cite the classical and pioneer books in the field (see Akhiezer (1965), Shohat and Tamarkin (1943)) which put emphasis on the existence aspect of the solution and its uniqueness. According to the support of the distribution of interest, one may refer to one of the three types of classical moment problems: the Hamburger moment problem whose support of μ is the whole real line, the Stieljes problem on $[0, +\infty)$ and finally the Hausdorff problem on a bounded interval. In this paper, we shall focus on the last issue under the inverse problem angle.

The Hausdorff moment problem which dates back to 1921 (see Hausdorff (1921)) occupies a central place in the field of inverse problems and has been an object of great interest in literature since then. For instance, the particular case when only a finite number of moments are known which is called the Truncated Hausdorff moment problem, has recently aroused much attention (see Talenti (1987), Fasino and Inglese (1996), Inglese (1995), Tagliani (2002)). Another interest and aspect of the Hausdorff moment problem lies in its very close link to the inversion of the Laplace transform when this latter is given at equidistant points (see for instance Brianzi and Frontini (1991), Dung, Huy, Quan and Trong (2006), Tagliani (2001), Tagliani and Velazquez (2003)). In fact, by a simple change of variable, the problem of Laplace transform inversion is equivalent to the Hausdorff moment problem. More recently, Ang, Gorenflo, Le and Trong (2002) have presented the Hausdorff moment problem under the angle of ill posed problems, in a sense that solutions do not depend continuously on the given data. Nonetheless, until now, as far as we know, the statistical approach which consists in assuming that the noise is stochastic has been very little put forward and rarely raised.

We consider in this paper a statistical point of view of the Hausdorff moment problem. We aim at reconstructing an unknown function from noisy measurements of its moments on a symmetric bounded interval in a statistical inverse problem framework. In reality, it is barely impossible to measure moments with having any corruption. Without loss of generality we may and will suppose that $[-a, a] = [-1, 1]$. In practise, observations of moments appear especially in quantum physics (see for instance Ash and Mc Donald (2003) and Mead and Papanicolaou (1983)), image analysis (see Teague (1980)), engineering mechanics (see Athanassoulis and Gavriliadis (2002)) and all the references therein). We shall give now some concrete examples of the application of the problem of moments, in which either one can have access directly to moments to reconstruct the unknown function at stake or the moment problem appears in the line of reasoning.

Examples. 1. In quantum physics, a big issue aims at reconstructing a positive density of states in harmonic solids from its moments on a finite interval. It is explained how to measure these moments (p 2410 and 2411 in Mead and Papan-

icolaou (1983)) and (p 3078 and 3079 in Gaspard and Cyrot-Lackmann (1973)).
2. In the field of quantum gravity, the transition probabilities for a Markov chain related to the causal set approach to modeling discrete theories of quantum gravity satisfy a moment problem (see Ash and McDonald (2003)). One has to measure those probabilities.
3. In the context of non classical moment problem, we may cite the following example (see Ang, Nhan and Thanh (1999)). It deals with the determination of the shape of an object in the interior of the Earth by gravimetric methods. The density of that object differs from the density of the surrounding medium. Assuming a flat earth model, the problem consists in finding a curve $x \to \sigma(x)$ in the half plane $0 \leq \sigma(x) < H, 0 \leq x \leq 1$, $\sigma(x)$ satisfying a non linear integral equation of the first kind of the form

$$\frac{1}{2\pi}\int_0^1 \frac{H-\sigma(\xi)}{(x-\xi)^2+(H-\sigma(\xi))^2}d\xi = f(x),$$

where $f(x)$ is a given function. The nonlinear integral equation can be approximated by the following linear integral equation in φ:

$$\int_0^1 \frac{\varphi(\xi)}{(M+x+\xi)^2}d\xi = 2\pi f(-M-x),$$

with $\varphi(x) = H-\sigma(x)$, M is large enough and $x \geq 0$. By taking $x = 1, 2, \ldots, n, \ldots$, we get the following equivalent moment problem:

$$\int_0^1 \frac{\varphi(\xi)}{(M+n+\xi)^2}d\xi = \mu_n,$$

where $\mu_n = 2\pi f(-M-n)$, $n = 1, 2, \ldots$

In the first examples cited above, the authors aim at estimating an unknown density from measurements of its moments. The following paper presents the case of the reconstruction of an unknown function which could be in particular a density of probability, when one has its moments corrupted with some white noise. This approach constitutes one among other ill-posed inverse problems point of view.

Estimation in statistics using moments has already been put forward (see Mnatsakanov and Ryumgaart (2005)), but the approach there was based on empirical moments and empirical processes, the results were expressed in terms of weak convergence whereas our paper is built upon an ill-posed inverse approach with minimax results.
The estimation procedure we use is based on the expansion of the unknown function through the basis of Legendre polynomials and an orthogonal series method. We establish an upper bound and a lower bound on the estimation accuracy of the procedure showing that it is optimal in a minimax sense. We show that the achieved rate is only of logarithmic order. This fact has already been underlined by Goldenshluger and Spokoiny (2004). In their paper, Goldenshluger and Spokoiny (2004) tackled the problem of reconstructing a planar convex set from noisy geometric moments observations. They pointed out that in view of reconstructing a planar region from noisy measurements of moments, the upper bound was only in the order of logarithmic rate. The lower bound has not been proved. In a second part, they consider reconstruction from Legendre moments to get faster rates of convergence. Legendre moments can be observed in the context of shapes reconstruction. In our present work, instead of considering a planar region, we deal with functions belonging to a Sobolev scale and we stay focused on the classic moments with respect to the monomials x^k. Moreover, recently in the context of long-memory processes obtained by aggregation of independant parameter AR(1) processes and in view of estimating the density of the underlying random parameter, Leipus, Oppenheim, Philippe and Viano (2006) had to deal with a problem of moments. They obtained very slow logarithmic rate but without showing that this could be the best possible. In a certain way, our minimax results provide a piece of answer.

One might question this chronic slow rate which seems inherent to moment problems. In fact, the underlying problem lies in the non orthogonal nature of the monomials x^k. They actually hamper the convergence rate to be improved for bringing a small amount of information. This remark is highlighted in our proof of the upper bound.

This paper is organized as follows: in section 2 we introduce the model and the estimator of the unknown function and we finally state the two theorems.

Section 3 contains the proofs. The last section is an appendix in which we prove some useful inequalities about binomial coefficients.

2. Statement of the problem

2.1 The model. First of all, let us recall an usual statistical framework of ill-posed inverse problems (see Mathe and Pereverzev (2001)):
Let $A : H \longrightarrow H$ be a known linear operator on a Hilbert space H. The problem is to estimate an unknown function $f \in H$ from indirect observations

$$Y = Af + \varepsilon\xi, \tag{1.1}$$

where ε is the amplitude of the noise. It is supposed to be a small positive parameter which tends to 0, ξ is assumed to be a zero-mean Gaussian random process indexed by H on a probability space $(\Omega, \mathcal{A}, \mathbf{P})$.
In the Hausdorff moment problem, the operator A_k which determines the way the observations are indirect is defined by:

$$A_k(f) = \int_{-1}^{1} x^k f(x)dx, \quad k \in \mathbb{N}. \tag{1.2}$$

Let us now state the Hausdorff moment problem model. From the equation (1.1) and (1.2), we derive the following sequence of moments observations pertubated by a stochastic noise:

$$y_k = \mu_k + \varepsilon\xi_k \quad k = 0, 1, \ldots. \tag{1.3}$$

where ε is the noise level, it is supposed to tend to 0, ξ_k are assumed to be i.i.d standard Gaussian random variables and μ_k are the moments of the unknown function f given by:

$$\mu_k = \int_{-1}^{1} x^k f(x)dx \quad k = 0, 1, \ldots.$$

The assumption that the noise is modelled as i.i.d standard Gaussian random variables is a natural modeling and has already been used (see Rodriguez and Seatzu (1993), Brianzi and Frontini (1991), Goldenshluger and Spokoiny (2004)). It is a standard assumption in statistical inverse problems. Moreover, we may

consider (1.3) as a first approach of the problem at hand.
The objective is to estimate the unknown function f supported on the bounded interval $[-1,1]$ from noisy observations of its moments in the model (1.3) which can be assimilated to a gaussian white noise model but using a non orthonormal basis and which constitutes an inverse problem setting. The solution of this problem is unique in the case $k=\infty$ (see Bertero, De Mol and Pike (1985)), so that the statistical problem of recovering f from (1.3) is relevant. In the particular case of estimating a density of probability, the probability measure having density f is unique (see Feller (1968, Chap.7)).
The use of the Legendre polynomials in the Hausdorff classical moment problem in order to approximate the unknown measure is a completely standard and quite natural procedure (see Ang, Gorenflo, Le and Trong (2002), Bertero, De Mol and Pike (1985), Papoulis (1956), Teague (1980), Goldenshluger and Spokoiny (2004)) as those polynomials directly result from the Gram-Schmidt orthonormalization of the family $\{x^k\}$, $k=0,1,\ldots$ Consequently, expanding the function f to be estimated in the basis of Legendre polynomials falls naturally and fits the problem's nature.
Denote $\beta_{n,j}$ the coefficients of the normalized Legendre polynomial of degree n:

$$P_n(x)=\sum_{j=0}^{n}\beta_{n,j}\,x^j$$

By considering the Legendre polynomials, if we multiply both sides of the model (1.3) by the coefficients $\beta_{k,j}$ we get the following model (for the proof see Lemma 3. in Appendix):

$$\tilde{y}_k=\theta_k+\varepsilon\sigma_k\xi_k \tag{1.4}$$

where $\sigma_k^2=\sum_{j=0}^{k}\beta_{k,j}^2$, $\tilde{y}_k=\sum_{j=0}^{k}\beta_{k,j}y_j$, $\theta_k=\sum_{j=0}^{k}\beta_{k,j}\mu_j=\int_{-1}^{1}f(x)P_k(x)dx$ and ξ_k are i.i.d standard Gaussian random variables. So the model (1.3) is equivalent to the model (1.4). The model (1.4) will be used for the proof of the lower bound.
Before going any further, we can make a remark at this stage concerning the model (1.4) which is an heteroscedastic gaussian sequence space model. Depending on the asymptotic behavior of the intensity noise σ_n^2 one may characterize the nature of the problem's ill-posedness (see Cavalier, Golubev, Lepski and Tsy-

bakov (2004)). Here, in our case, $\sigma_n^2 \geq \frac{1}{4}4^n$ (see Lemma 5. from Appendix) and hence tends to infinity exponentially. We may say that we are dealing with a severely ill-posed problem with log-rates .

We assume that f belongs to the Sobolev space W_2^r defined by:

$$W_2^r = \{f \in L^2[-1,1] : \sum_k k^{2r}|\theta_k|^2 < \infty\}$$

where $\theta_k = \int_{-1}^1 f(x)P_k(x)dx$ is the Legendre Fourier coefficient and P_k denotes the normalized Legendre polynomial of degree k. Note that we consider more general functions than densities of probability which are included in this smoothness class.

Sobolev spaces associated with various kinds of underlying orthonormal basis constitute quite standard smoothness assumption classes in classical ill-posed problems (see for instance Mair and Ruymgaart (1996), Mathe and Pereverzev (2002), Goldenshluger and Pereverzev (2000)). In the Hausdorff moment problem, the underlying basis is the Legendre polynomials.

Let us now give some highlights of the Sobolev space W_2^r regarding Legendre polynomials. Rafal'son (1968) and Tomin (1973) have shown in the more general case of Jacobi polynomials (and thus in the particular case of Legendre polynomials we are considering here) that Sobolev space W_2^r consists of all functions f which have their derivatives $f', f'', \ldots, f^{(r-1)}$ being absolutely continuous on each interval $[a,b] \subset (-1,1)$ (see also Mathe and Pereverzev (2002)).

2.2 The estimation procedure. Let us define now the estimator of f. This latter is induced by an orthogonal series method through the Legendre polynomials.

Any function in $L^2[-1,1]$ has an expansion:

$$f(x) = \sum_{k=0}^{\infty} \theta_k P_k(x) \quad \text{with} \quad \theta_k = \int_{-1}^{1} f(x)P_k(x)dx.$$

The problem of estimating f reduces to estimation of the sequence $\{\theta_k\}_{k=1}^{+\infty}$ for Legendre polynomials form a complete orthogonal function system in $L^2[-1,1]$.

We have

$$\theta_k = \int_{-1}^{1} f(x)P_k(x)dx = \sum_{j=0}^{k} \beta_{k,j} \int_{-1}^{1} f(x)x^j dx = \sum_{j=0}^{k} \beta_{k,j}\mu_j.$$

This leads us to consider the following estimator of θ:

$$\hat{\theta}_k = \sum_{j=0}^{k} \beta_{k,j} y_j$$

and hence the estimator $\hat{f}_N$ of f:

$$\hat{f}_N(x) = \sum_{k=0}^{N} \hat{\theta}_k P_k(x) = \sum_{k=0}^{N}\sum_{j=0}^{k} \beta_{k,j} y_j P_k(x)$$

where y_j is given by (1.3) and N is an integer to be properly selected later.

The mean integrated square error of the estimator $\hat{f}_N$ is:

$$\mathbb{E}_f \|\hat{f}_N - f\|^2,$$

where $\mathbb{E}_f$ denotes the expectation w.r.t the distribution of the data in the model (1.3) and for a function $g \in L^2[-1,1]$,

$$\|g\| = \left(\int_{-1}^{1} g^2(x)dx \right)^{1/2}.$$

In this paper we shall consider the problem of estimating f using the mean integrated square risk in the model (1.3).

We state now the two results of the paper. The first theorem establishes an upper bound.

Theorem 1. *For $\alpha > 0$, define the integer $N = \lfloor \alpha \log(1/\varepsilon) \rfloor$. Then we have*

$$\sup_{f \in \mathbb{W}_2^r} \mathbb{E}_f \|\hat{f}_N - f\|^2 \leq C[\log(1/\varepsilon)]^{-2r},$$

where C is an absolute positive constant and $\lfloor \cdot \rfloor$ denotes the floor function.

We recall that the floor function is defined by: $\lfloor x \rfloor = \max\{n \in \mathbb{Z} | \, n \leq x\}$.
The second theorem provides a lower bound.

Theorem 2. *We have*

$$\inf_{\hat{f}} \sup_{f\in\mathbb{W}_2^r} \mathbb{E}_f \|\hat{f} - f\|^2 \geq c[\log(1/\varepsilon)]^{-2r},$$

where c is a positive constant which depends only on r and the infimum is taken over all estimators $\hat{f}$.

3. Proofs

3.1 Proof of Theorem 1. For the following proof, we consider the genuine model (1.3). By the usual MISE decomposition which involves a variance term and a bias term, we get

$$\mathbb{E}_f\|\hat{f}_N - f\|^2 = \mathbb{E}_f \sum_{k=0}^{N} (\hat{\theta}_k - \theta_k)^2 + \sum_{k\geq N+1} \theta_k^2$$

but

$$\begin{aligned}\mathbb{E}_f \sum_{k=0}^{N} (\hat{\theta}_k - \theta_k)^2 &= \mathbb{E}_f \sum_{k=0}^{N} (\sum_{j=0}^{k} \beta_{k,j}(y_j - \mu_j))^2 \\ &= \varepsilon^2 \mathbb{E}_f \sum_{k=0}^{N} (\sum_{j=0}^{k} \beta_{k,j}\xi_j)^2\end{aligned}$$

and since $\xi_j \overset{iid}{\sim} N(0,1)$, it follows that

$$\begin{aligned}\mathbb{E}_f\|\hat{f}_N - f\|^2 &= \varepsilon^2 \sum_{k=0}^{N}\sum_{j=0}^{k} \beta_{k,j}^2 + \sum_{k\geq N+1} \theta_k^2 \\ &= V_N + B_N^2\end{aligned}$$

We first deal with the variance term V_N. To this end, we have to upper bound the sum of the squared coefficients of the normalized Legendre polynomial of degree k. Set $\sigma_k^2 = \sum_{j=0}^{k} \beta_{k,j}^2$. An explicit form of $P_k(x)$ is given by (see Abramowitz and Stegun (1970)):

$$P_k(x) = \left(\frac{2k+1}{2}\right)^{1/2} \frac{1}{2^k} \sum_{j=0}^{[k/2]} (-1)^k \binom{k}{j}\binom{2k-2j}{k} x^{k-2j},$$

where $[\cdot]$ denotes the integer part and $\binom{k}{j}$ denotes the binomial coefficient, $\binom{k}{j} = \frac{k!}{(k-j)!j!}$. This involves

$$
\begin{aligned}
\sigma_k^2 &= \frac{2k+1}{2}\frac{1}{4^k}\sum_{j=0}^{[k/2]}\left\{\binom{k}{j}\binom{2k-2j}{k}\right\}^2 \\
&\le \frac{2k+1}{2}\frac{1}{4^k}\left\{\binom{2k}{k}\right\}^2\sum_{j=0}^{[k/2]}\left\{\binom{k}{j}\right\}^2 \qquad (3.1)\\
&\le \frac{2k+1}{2}\frac{1}{4^k}\left\{\binom{2k}{k}\right\}^2(2^k)^2, \qquad (3.2)
\end{aligned}
$$

the inequality (3.1) is due to the fact that for $0 \le j \le [k/2]$, we have $\binom{2k-2j}{k} \le \binom{2k}{k}$. As for (3.2), we have $\sum_{j=0}^{[k/2]}\left\{\binom{k}{j}\right\}^2 \le \left\{\sum_{j=0}^{k}\binom{k}{j}\right\}^2$ and it is well known that $\left\{\sum_{j=0}^{k}\binom{k}{j}\right\}^2 = (2^k)^2$.
By using now that $\{\binom{2k}{k}\}^2 \le \frac{4^{2k}}{\sqrt{k}}$ (see Lemma 4. from Appendix) we have

$$\sigma_k^2 \le \frac{2k+1}{2}\frac{4^{2k}}{\sqrt{k}}$$

which yields

$$V_N \le C\varepsilon^2 N^{3/2}4^{2N},$$

where $C > 0$ denotes an absolute positive constant.

Now, it remains to upper bound the bias term B_N^2.

$$
\begin{aligned}
B_N^2 &= \sum_{k\ge N+1}\theta_k^2 \\
&= \sum_{k\ge N+1}\frac{k^{2r}}{k^{2r}}\theta_k^2 \\
&\le N^{-2r}\sum_{k=1}^{\infty}k^{2r}\theta_k^2
\end{aligned}
$$

Since the function f belongs to the space W_2^r, $\sum_k k^{2r}|\theta_k|^2 < \infty$, we get

$$B_N^2 = \mathcal{O}(N^{-2r})$$

Finally we have the upper bound for the MISE:

$$\mathbb{E}_f\|\hat{f}_N - f\|^2 \le C\varepsilon^2 N^{3/2}4^{2N} + C'N^{-2r} \qquad (3.3)$$

At last, it remains to choose the optimal N which will minimize the expression (3.3). This N is obtained by equalizing the upper bounds of the bias and the variance term, namely:

$$C\varepsilon^2 N^{3/2} 4^{2N} = C' N^{-2r},$$

as $4^{2N} \gg N^{2r+3/2}$, consequently $N \asymp \log(\frac{1}{\varepsilon^2})$. Once one plugs $N \asymp \log(\frac{1}{\varepsilon^2})$ in (3.3), the desired result of the Theorem 1. follows.

3.2 Proof of Theorem 2. From now on, to prove the lower bound and for practical reasons, we shall consider the model (1.4) which constitutes a heteroscedastic gaussian sequence space model. We recall that the equivalence between the models (1.3) and (1.4) is proved in Lemma 3. in Appendix.
A successful approach and standard tool to obtain lower bounds for minimax risk consists in specifying a subproblem namely constructing a subset of functions based on the observations (1.4). Then we lean on the application of the following particular version of Fano's lemma (see Birgé and Massart (2001)) which will allow us to evaluate the difficulty of the specified subproblem and will give us a lower bound for the MISE associated to this subproblem.
One crucial point in the Fano's lemma is the use of the Kullback-Leibler divergence $K(\mathbb{P}_1, \mathbb{P}_0)$ between two probability distributions $\mathbb{P}_1$ and $\mathbb{P}_0$ defined by:

$$K(\mathbb{P}_1, \mathbb{P}_0) = \begin{cases} \int_{\mathbb{R}} \log(\frac{p_1(x)}{p_0(x)}) p_1(x) dx & \text{if } \mathbb{P}_1 \ll \mathbb{P}_0 \\ +\infty & \text{otherwise} \end{cases}$$

Here's the version of Fano's lemma, we are going to exploit:

Lemma 1. *Let η be a strictly positive real number, $\mathcal{C}$ be a finite set of elements $\{f_0, \dots, f_M\}$ on $\mathbb{R}$ with $|\mathcal{C}| \geq 6$ and $\{P_j\}_{j \in \mathcal{C}}$ a set of probability measures indexed by $\mathcal{C}$ such that :*
(i) $\|f_i - f_j\| \geq \eta > 0, \quad \forall\, 0 \leq i < j \leq M$.
(ii) $\mathbb{P}_j \ll \mathbb{P}_0, \quad \forall\, j = 1, \dots, M$, and

$$K(\mathbb{P}_j, \mathbb{P}_0) \leq H < \log M$$

then for any estimator $\hat{f}$ and any nondecreasing function ℓ

$$\sup_{f \in \mathcal{C}} \mathbb{E}_f \big[\ell(\|\hat{f} - f\|)\big] \geq \ell(\frac{\eta}{2}) \Big[1 - \Big(\frac{2}{3} \vee \frac{H}{\log M}\Big)\Big].$$

First of all, we have to construct an appropriate set of functions $\mathcal{E}$. We are going to define $\mathcal{E}$ as a set of functions of the following type

$$\mathcal{E} = \Big\{ f_\delta \in W_2^r : f_\delta = \mathbf{1}_{[-1,1]}\Big(\frac{c_0}{m^{(4r+3)/2}} \sum_{k=m}^{2m-1} \delta_k k^{(2r+2)/2} P_k\Big), \\ \delta = (\delta_m, \dots, \delta_{2m-1}) \in \Delta = \{0,1\}^m \Big\}$$

We verify that f_δ belongs to W_2^r. In this aim, we have to calculate the Legendre-Fourier coefficients associated with the function f_δ:

$$\begin{aligned} \theta_{\delta\, l} &= \int_{-1}^{1} f_\delta(x) P_l(x) dx \\ &= \begin{cases} \frac{c_0}{m^{(4r+3)/2}} \cdot l^{(2r+2)/2} \cdot \delta_l & \text{if } l \in [m, 2m-1] \\ 0 & \text{else} \end{cases} \end{aligned} \tag{3.4}$$

hence

$$\begin{aligned} \sum_{k=0}^{+\infty} k^{2r} \theta_{\delta\, k}^2 &= \frac{c_0^2}{m^{4r+3}} \sum_{k=m}^{2m-1} k^{2r} k^{2r+2} \delta_k^2 \\ &\leq \frac{c_0^2}{m^{4r+3}} \sum_{k=m}^{2m-1} k^{4r+2} \delta_k^2 \\ &\leq \frac{c_0^2 (2m)^{4r+2}}{m^{4r+3}} \sum_{k=m}^{2m-1} \delta_k^2 \leq c_0^2 2^{4r+2} < \infty, \end{aligned}$$

since $\delta_k \in \{0,1\}$.

We set $\delta^{(0)} = (0, \dots, 0)$ and $f_{\delta^{(0)}} \equiv f_0$. The Legendre-Fourier coefficients of f_0 are null:

$$\theta_{0\, l} = 0 \quad \forall\, l \in \mathbb{N}. \tag{3.5}$$

We are now going to exhibit the suitable subset of functions $\mathcal{C}$ of the Lemma 1. To this purpose, we only take into consideration a subset of $M+1$ functions of $\mathcal{E}$:

$$\mathcal{C} = \{f_{\delta^{(0)}}, \dots, f_{\delta^{(M)}}\}$$

where $\{\delta^{(1)},\dots,\delta^{(M)}\}$ is a subset of $\{0,1\}^m$.
We precise that $\mathbb{P}_\delta$ is the law of the vector of observations $\tilde{Y}=(\tilde{y}_1,\dots,\tilde{y}_\infty)$ in the model (1.4) for $f=f_\delta$, $\delta\in\mathcal{C}$.

We are now going to apply Lemma 1. We first check the condition (i), accordingly, we have to assess the distance $\|f_{\delta^{(i)}}-f_{\delta^{(j)}}\|^2$. By the orthogonality of the system $\{P_k\}_k$ and thanks to Parseval equality we get, for $0\le i<j\le M$,

$$\begin{aligned}\|f_{\delta^{(i)}}-f_{\delta^{(j)}}\|^2 &= \frac{c_0^2}{m^{4r+3}}\sum_{k=m}^{2m-1}k^{2r+2}(\delta_k^{(i)}-\delta_k^{(j)})^2\\ &\ge \frac{c_0^2}{m^{4r+3}}\cdot m^{2r+2}\sum_{k=m}^{2m-1}(\delta_k^{(i)}-\delta_k^{(j)})^2\\ &\ge \frac{c_0^2}{m^{2r+1}}\sum_{k=m}^{2m-1}(\delta_k^{(i)}-\delta_k^{(j)})^2\\ &= \frac{c_0^2}{m^{2r+1}}\rho(\delta^{(i)},\delta^{(j)}),\end{aligned}$$

where $\rho(\cdot,\cdot)$ is the *Hamming distance*. We are going to resort to the Varshamov-Gilbert bound which is stated in the following lemma to find a lower bound of the quantity $\rho(\delta^{(i)},\delta^{(j)})$:

Lemma 2. (Varshamov-Gilbert bound, 1962). *Fix $m\ge 8$. Then there exists a subset $\{\delta^{(0)},\dots,\delta^{(M)}\}$ of Δ such that $M\ge 2^{m/8}$ and*

$$\rho(\delta^{(j)},\delta^{(k)})\ge\frac{m}{8},\quad \forall\, 0\le j<k\le M.$$

Moreover we can always take $\delta^{(0)}=(0,\dots,0)$.

For a proof of this lemma see for instance Tsybakov (2004), p 89.
Hence

$$\|f_{\delta^{(i)}}-f_{\delta^{(j)}}\|^2\ge (c_0^2)/(8m^{2r})\equiv\eta^2$$

We are now going to check the condition (ii) in Lemma 1. and evaluate the Kullback-Leibler divergence. It is well known (see for instance Birgé and Massart (2001) p 62) that for the Kullback-Leibler divergence in the case of a gaussian sequence space model we have

$$K(\mathbb{P}_\delta,\mathbb{P}_0) = \frac{1}{\varepsilon^2}\sum_{l=1}^{\infty}\frac{|\theta_{\delta\, l}-\theta_{0\, l}|^2}{\sigma_l^2}. \tag{3.6}$$

Hence, by virtue of (3.4), (3.5) and (3.6), the Kullback-Leibler divergence between the two probability measures $\mathbb{P}_0$ and $\mathbb{P}_\delta$ of the observations in the model (1.4) associated respectively with functions $f = f_0$ anf $f = f_\delta$ for all $\delta \in \mathcal{C}$ satisfies

$$\begin{aligned} K(\mathbb{P}_\delta, \mathbb{P}_0) &= \frac{1}{\varepsilon^2} \frac{c_0^2}{m^{4r+3}} \sum_{l=m}^{2m-1} \frac{l^{2r+2}\delta_l^2}{\sigma_l^2} \\ &\leq \frac{c_0^2 2^{2r+2}}{\varepsilon^2} \frac{m^{2r+2}}{m^{4r+3}} \sum_{l=m}^{2m-1} \frac{\delta_l^2}{\sigma_l^2} \end{aligned}$$

but thanks to Lemma 5. (see Appendix) we have

$$\frac{1}{\sigma_l^2} \leq \frac{1}{4^{l-1}}$$

which implies

$$K(\mathbb{P}_\delta, \mathbb{P}_0) \leq \frac{c_0^2 2^{2r+4}}{\varepsilon^2} \frac{1}{m^{2r+1} 4^m} \sum_{l=m}^{2m-1} \delta_l^2 \leq \frac{c_0^2 2^{2r+4}}{\varepsilon^2} \frac{1}{m^{2r+1} 4^m} \cdot m \leq \frac{c_0^2 2^{2r+4} m}{\varepsilon^2 4^m}$$

One chooses $m = \frac{1}{\log 4} \log(\frac{1}{\varepsilon^2})$ so that $1/4^m = \varepsilon^2$, hence:

$$K(\mathbb{P}_\delta, \mathbb{P}_0) \leq c_0^2 2^{2r+4} m$$

and since $m \leq 8 \log M / \log 2$ (see Lemma 2)

$$K(\mathbb{P}_\delta, \mathbb{P}_0) \leq \frac{c_0^2 2^{2r+7}}{\log 2} \log M$$

Eventually one can choose c_0 small enough to have $c \equiv \frac{c_0^2 2^{2r+7}}{\log 2} < 1$.

Since now all the conditions of the Lemma 1. are fulfilled, we are in position to apply its result with the loss function $\ell(x) = x^2$ and $\eta = (c_0)/(2\sqrt{2} m^r)$. Therefore, we derive that whatever the estimator $\hat{f}$,

$$\sup_{f \in \mathcal{C}} \mathbb{E}_f[\|\hat{f} - f\|^2] \geq \frac{c_0^2}{32 m^{2r}} \left[1 - \left(\frac{2}{3} \vee \frac{c}{\log M} \right) \right] \geq \frac{c_0^2}{96 m^{2r}}.$$

but from above we had $m = \frac{1}{\log 4} \log(\frac{1}{\varepsilon^2})$ which gives the desired lower bound.

4. Conclusion

The two theorems of this paper show that in the problem of estimating a function on a compact interval from noisy moments observations, the best rate of convergence one can achieve, supposing Sobolev scale smoothness and considering the mean integrated squared error, is only of logarithmic order. In a future work, one could try to generalise this result for L^p loss and may obtain faster rate of convergence if one assumes a more restricted smoothness class involving super smooth functions. Besides, one may consider an heteroscedastic gaussian noise instead of a white noise model.

5. Appendix

Lemma 3. *The models (1.3) and (1.4) are equivalent.*

Proof. We recall the model (1.3):

$$y_j = \mu_j + \varepsilon \xi_j = \int_{-1}^{1} f(x) x^j dx + \varepsilon \xi_j.$$

We are going now to multiply both sides of (1.3) by the coefficient β_{jk} of the Legendre polynomial:

$$\begin{aligned}
(1.3) \quad \Leftrightarrow \quad & \beta_{kj} y_j = \beta_{kj} \int_{-1}^{1} f(x) x^j dx + \varepsilon \beta_{kj} \xi_j \\
\Leftrightarrow \quad & \sum_{j=0}^{k} \beta_{kj} y_j = \sum_{j=0}^{k} \beta_{kj} \int_{-1}^{1} f(x) x^j dx + \sum_{j=0}^{k} \varepsilon \beta_{kj} \xi_j \\
\Leftrightarrow \quad & \sum_{j=0}^{k} \beta_{kj} y_j = \int_{-1}^{1} f(x) \sum_{j=0}^{k} \beta_{kj} x^j dx + \varepsilon \sum_{j=0}^{k} \beta_{kj} \xi_j \\
\Leftrightarrow \quad & \sum_{j=0}^{k} \beta_{kj} y_j = \int_{-1}^{1} f(x) P_k dx + \varepsilon \sum_{j=0}^{k} \beta_{kj} \xi_j
\end{aligned}$$

Let us set $\tilde{\xi}_k = \sum_{j=0}^{k} \beta_{kj} \xi_j$. Since ξ_j are i.i.d standard Gaussian random variables, the random variable $\tilde{\xi}_k$ follows a normal law with zero mean and variance equal to $\sum_{j=0}^{k} \beta_{kj}^2$. Hence (1.3) is equivalent to:

$$\tilde{y}_k = \theta_k + \varepsilon \sigma_k \xi_k$$

where $\sigma_k^2 = \sum_{j=0}^k \beta_{k,j}^2$, $\tilde{y}_k = \sum_{j=0}^k \beta_{k,j} y_j$, $\theta_k = \sum_{j=0}^k \beta_{k,j}\mu_j = \int_{-1}^1 f(x)P_k(x)dx$ and ξ_k are i.i.d standard Gaussian random variables.

Lemma 4. *For all $n \geq 1$ we have:*

$$\binom{2n}{n} \leq \frac{4^n}{n^{1/4}} \tag{5.1}$$

Proof. Let us prove (5.1) by recursion on n. The inequality is clearly true for $n = 1$.

Suppose (5.1) true for a certain $n \geq 1$.

$$\binom{2(n+1)}{n+1} = \binom{2n}{n}\frac{2(2n+1)}{n+1} \leq \frac{4^n}{n^{1/4}}\frac{2(2n+1)}{n+1},$$

by recursion hypothesis. It remains to prove that

$$\frac{4^n}{n^{1/4}}\frac{2(2n+1)}{n+1} \leq \frac{4^{n+1}}{(n+1)^{1/4}}. \tag{5.2}$$

$$\begin{aligned}
(5.2) \quad &\Longleftrightarrow \quad \frac{2(2n+1)}{n^{1/4}(n+1)} \leq \frac{4}{(n+1)^{1/4}} \\
&\Longleftrightarrow \quad \frac{n+1}{n}\left(\frac{2n+1}{n+1}\right)^4 \leq 2^4 \\
&\Longleftrightarrow \quad (n+\frac{1}{2})^{1/4} \leq n(n+1)^3,
\end{aligned}$$

which is true because we have $(n+\frac{1}{2})^{1/4} \leq (n+\frac{1}{2})^3(n+1)$ and $(n+\frac{1}{2})^3 \leq n(n+1)^2$ since $\frac{1}{8} \leq n^2/2 + n/4$. This completes the proof.

Lemma 5. *For all $n \geq 1$, we have:*

$$\sigma_n^2 \geq 4^{n-1} \tag{5.3}$$

where σ_n is defined in (1.4).

Proof. Firstly, let us recall the value of the noise intensity σ_n^2:

$$\begin{aligned}
\sigma_n^2 &= \frac{2n+1}{2}\frac{1}{4^n}\sum_{j=0}^{[n/2]}\left\{\binom{n}{j}\binom{2n-2j}{n}\right\}^2 \\
&\geq \frac{n}{4^n}\binom{2n}{n}^2.
\end{aligned}$$

And so, in order to prove (5.3) it remains to prove that

$$\binom{2n}{n} \geq \frac{4^n}{2\sqrt{n}} \qquad n \geq 1.$$

We again use a recursion on n.
The inequality (5.3) is clear for $n = 1$. We suppose the property true for a certain $n \geq 1$ and we shall prove it at the rank $(n+1)$.

$$\begin{aligned} \binom{2(n+1)}{n+1} &= \binom{2n}{n} \frac{2(2n+1)}{n+1} \\ &\geq \frac{4^n}{2\sqrt{n}} \frac{2(2n+1)}{n+1} \\ &> \frac{4^{n+1}}{2\sqrt{n+1}} \end{aligned} \tag{5.4}$$

the inequality (5.4) is true because it is equivalent to $4n^2 + 4n + 1 > 4n^2 + 4n$ what we always have.

Acknowledgements

The author wishes to thank the referees for their valuable criticisms and suggestions and Dominique Picard for interesting discussions, leading to the present improved version of the paper.

References

Abramowitz, M. and Stegun, I. A. (1970). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical tables*. Dover Publications, New York.

Akhiezer, N. I. (1965). *The classical Moment Problem and Some Related Questions in Analysis*. Oliver and Boyd, Edinburgh.

Ang, D. D., Gorenflo, R., Le, V. K. and Trong, D. D. (2002). Moment Theory and Some inverse Problems in Potential Theory and Heat Conduction. *Lectures Notes in Mathematics* **1792**, Springer, Berlin.

Ang, D. D., Nhan, V. N.; Thanh, D. N. (1999). A nonlinear integral equation of gravimetry: Uniqueness and approximation by linear moments. *Vietnam J. Math.*, **27**(1), 61-67. (1999).

Athanassoulis G.A. and Gavriliadis P. N. (2002). The truncated Hausdorff moment problem solved by using kernel density functions. *Probabilistic Engineering Mechanics*, **17**, 273-291.

Ash, A. and Mc Donald P. (2003). Moments problem and the causal set approach to quantum gravity. *Journal of Mathematical Physics*, **44**(4), 1666-1678.

Birgé, L. and Massart, P. (2001). Gaussian model selection. *J. Eur. Math. Soc*, **3**, 203-268.

Bertero, M., De Mol, C., and Pike, E. R. (1985). Linear inverse problems with discrete data. I: General formulation and singular system analysis. *Inverse Problems*, **1**, 301-330.

Brianzi, P. and Frontini, M. (1991). On the regulized inversion of the Laplace transform. *Inverse Problems*, **7**, 355-368.

Cavalier, L., Golubev, Y., Lepski, O., and Tsybakov, A. B. (2004). Block thresholding and sharp adaptative estimation in severely ill-posed inverse problems. *Theory Probab. Appl.*, **48**(3), 426-446.

Dung, N., Huy, N. V., Quan, P. H., Trong, D. D. (2006). A Hausdorff-like moment problem and the inversion of the Laplace transform. *Mathematische Nachrichten*, **279**(11), 1147-1158.

Fasino, D. and Inglese, G. (1996). Recovering a probability density from a finite number of moments and local a priori information. *Rend. Instituto. Mat. Univ. Trieste*, **28**(1), 184-200.

Feller, W. (1968). *An introduction to Probability Theory and its Applications. Vol II. 2nd ed.* Wiley Series in Probability and Mathematical Statistics, New York.

Gaspard, J.P. and Cyrot-Lackman, F. (1973). Density of states from moments. Application to the impurity band. *J. Phys. C: Solid State Phys*, **6**, 3077-3096.

Goldenshluger, A. and Pereverzev, S. (2000). Adaptive estimation of linear functionals in Hilbert scales from indirect white noise observations. *Probab. Theory Relat. Fields*, **118**, 169-186.

Goldenshluger, A. and Spokoiny, V. (2004). On the shape-from-moments problem and recovering edges from noisy Radon data. *Probab. Theory Relat. Fields*, **128**, 123-140.

Hausdorff, F. (1921). Summationsmethoden und Momentfolgen, I. *Mathematische Zeitschrift*, **9**, 74-109.

Inglese, G. (1995). Christoffel functions and finite moments problems. *Inverse problems*, **11**, 949-960.

Kay, M. and Marple, S. L. (1981). Spectrum analysis-a modern perspective. *Proc. IEEE*, **69**, 1380-1419.

Lang, S. W. and McClellan, J. H. (1983). Spectral estimation for sensor arrays. *IEEE Trans. Acoust. Speech Signal Process*, **31**, 349-358.

Leipus, R., Oppenheim, G., Philippe, A., Viano, M. C. (2006). Orthogonal series density estimation in a disaggregation scheme. *J. Stat. Plan. Inf.*, **138**(8), 2547-2571.

Mair B., Ruymgaart, F. H. (1996). Statistical estimation in Hilbert scales. *SIAM. J. Appl. Math.*, **56**, 1424-1444.

Mathe, P. and Pereverzev, S. V. (2001). Optimal discretization of inverse problems in Hilbert scales. Regularization and self-regularization of projections methods. *SIAM J. Numer. Anal.* **38**(6), 1999-2021.

Mathe, P. and Pereverzev, S. V. (2002). Stable summation of orthogonal series with noisy coefficients. *J. Approx. Theory* **118**(1), 66-80.

Mead, L. R. and Papanicolaou, N. (1984). Maximum entropy in the problem of moments. *J. Math. Phys.* **25**(8), 2404-2417.

Papoulis, A. (1956). A new method of inversion of the Laplace transform. *Q. Appl. Math.*, **14**, 405-414.

Rafal'son S.Z. (1968). The approximation of functions by Fourier-Jacobi sums. *Izv. Vyss. Ucebn. Zaved. Matematika* , **4**, 54-62.

Rodriguez, G. and Seatzu S. (1992). On the solution of the finite moment problem. *J. Math. Anal. Appl.*, **171** 321-333.

Rodriguez, G. and Seatzu S. (1993). Approximation methods for finite moment problem. *Numerical algorithms*, **5**(8) 391-405.

Shohat, J. A. and Tamarkin, J. D. (1943). *The problem of moments.* American Mathematical Society, New York.

Tagliani, A. (2001). Numerical inversion of Laplace transform on the real line of probability density functions. *Applied mathematics and computation*, **123**(3), 285-299.

Tagliani, A. (2002). Entropy estimate of probability densities having assigned moments: Hausdorff case. *Appl. Math. Lett.*, **15**(3), 309-314.

Tagliani, A. and Velazquez Y. (2003). Numerical inversion of the Laplace transform via fractional moments. *Applied mathematics and computation*, **143**(1),

Talenti, G. (1987). Recovering a function from a finite number of moments. *Inverse Problems*, **3**, 501-517.

Teague, M. R. (1980). Image analysis via the general theory of moments. *J. Opt. Soc. Am.*, **70**, 920-924.

Tomin, N. G (1973). An application of the interpolation of linear operators to questions of the convergence of series of Fourier coefficients with respect to classical orthogonal polynomials. *Dokl. Akad. Nauk. SSSR*, **212**, 1074-1077.

Tsybakov, A. B. (2004). *Introduction à l'estimation non-paramétrique.* Springer-Verlag. Berlin.

Laboratoire de Probabilités et Modèles aléatoires, UMR 7599, Université Paris 6, case 188, 4, Pl. Jussieu, F-75252 Paris Cedex 5, France.
E-mail: thanh.pham_ngoc@upmc.fr