L^2-DENSITY ESTIMATION UNDER CONSTRAINTS

Christian Musso[1], *Nadia Oudjane*[2]

[1] ONERA/DTIM, 29 avenue de la division Leclerc. 92322 Chatllon, France
[2] EDF R&D/OSIRIS, 1 av. du Général de Gaulle, 92140, Clamart, France
Email : christian.musso@onera.fr, nadia.oudjane@edf.fr

ABSTRACT

In this paper, we are interested in non parametric density estimation under constraints. It generalises a previous paper which was devoted to density estimation with non-positive kernels. The resulting density approximation improves the estimation (by reducing the bias) but provides negative values. Therefore, we have proposed a projection method on the space of probability densities and an algorithm designed to generate a sample from the projected density. We present here a generalization of this work in considering several linear constraints on the estimated density. These constraints represent an a priori knowledge of the underlying density. For example, the support, some moments or quantiles of the approximated density can be set a priori by the user. We prove that the projected density on the closed and convex set of functions satisfying some the constraints has a simple and explicit form. Some simulations show that the proposed solution outperforms alternative solutions proposed in the literature.

1. INTRODUCTION

In a previous paper we have proposed an optimal approximation of a function, which can take negative values, by a density. This kind of approximation can be useful when we use non-positive kernels for non-parametric density approximation. Let K be a kernel , f be a probability density defined on $\mathbb{R}^d$ and $(X_1,\cdots,X_N)$ be an i.i.d. sampled from f. The goal is to estimate f by $\hat{f}(x) = \frac{1}{Nh^d}\sum_{i=1}^N K(\frac{x-X_i}{h})$ [6]. In order to reduce the bias of the kernel approximation $\hat{f}$, one can use non-positive kernels ([2], [3], [7]). But, in that case, the approximation $\hat{f}$ can take negative values which is of course undesirable. In [10], we have proposed to project the non-positive approximation $\hat{f}$ on the space of probability densities of L^2, this projection being the new approximation of f. This method can be applied when we approximate a function which have some a priori properties. We wish that the approximation inherits these properties. This is the case for example when we develop a density with the Edgeworth serie. The resulting expansion can produce negative values. We present here a generalisation of this work in considering several linear constraints of the estimated density. These constraints represent an a priori knowledge of the underlying density. For example, the support, some moments or quantiles of the approximated density can be set a priori by the user. We prove that the projected density on the closed and convex set of the constraints has a simple and explicit form. Some simulations show that the proposed solution outperforms alternative solutions proposed in the literature.

2. GENERAL FRAMEWORK

2.1 Motivation

- Positivity constraint : as described in the introduction density estimation with a non positive kernel can produce negative values. The Edgeworth expansion of the density of a sum of iid variables can also [4] produce negative values.
- Support constraint : in some cases, one can be interested in forcing the estimate to have a predefined support $C \subset \mathbb{R}^d$. For example in [8] a weighted bootstrap method is described in order to construct a density estimator for salary data which are necessarily positive. Suppose we want to approximate a real valued function f defined on $\mathbb{R}^d$ such that $\int f = 1$ by a probability density with support $C \subset \mathbb{R}^d$. One possibility, if $f \in L^1$, is to use the following approximation,
$$f_1(x) = \frac{f^+(x)\mathbf{1}_C(x)}{\int_{\mathbb{R}^d} f^+(x)\mathbf{1}_C(x)dx}\,, \quad \text{for all} \quad x \in \mathbb{R}^d\,, \quad (1)$$
where $(a)^+ = \max(a,0)$, for any $a \in \mathbb{R}$ and $\mathbf{1}_C(x) = 1$ if $x \in C$, 0 if not. An alternative, if f is in L^2 , is to construct the projection of f on the subspace of L^2-probability densities with support C. This projection improves the estimation for the L^2 error as well as for the L^1 error [10].
- Quantile or moment constraint : suppose you have a density estimation $\hat{f}$ based on a sample. In some cases you want to find the optimal density g minimizing the L^2 distance to $\hat{f}$, for which the quantile ξ_α such that $P(X \geq \xi_\alpha) = \alpha \in (0,1)$ (where X follows g) is a priori given. You can also impose the mean of g.

Beyond these three different cases we propose in the paper an optimal (in some sense) density approximation method. Thereafter, results will be stated in the d-dimensional case.

2.2 General formulation

Let us consider an L^2 function f defined on $\mathbb{R}^d$ with values in $\mathbb{R}$. Let $f_1,\cdots,f_n$ be a collection of functions defined on $\mathbb{R}^d$ with values in $\mathbb{R}$ and $d_1,\cdots,d_n$ some reals. Let S be the convex set of L^2 defined as follows
$$S = \left\{ g \in L^2 \,|\, g \geq 0 \quad \text{and} \int g f_i = d_i \quad \text{for } i = 1\cdots n \right\}\,. \quad (2)$$
Thereafter we will assume that the constraint functions are such that S is closed. In this paper, we will be specifically interested in the case where one of the constraints is such that $f_1 \equiv 1$ and $d_1 = 1$ i.e. where S is a subspace of $L^2 \cap P(\mathbb{R}^d)$ with $P(\mathbb{R}^d)$ denoting the space of probability densities

defined on $\mathbb{R}^d$. The aim of this paper is to provide in most cases a simple expression for the projection g^* of f on the closed convex set S:

$$g^* = \arg\min_{g \in S} \int |f-g|^2 \ . \tag{3}$$

3. CHARACTERIZATION OF THE SOLUTION

In this section, we are interested in characterizing the general form of the projection (3). In the first subsection, we consider the projection as the solution of an optimization problem with linear constraints and a positivity constraint. The Lagrange method provides the solution which is of the form $g^* = (f - \sum_{i=1}^n \alpha_i^* f_i)^+$, in the special case where the constraint functions f_i belong to L^2. Unfortunately, this approach is not valid if we are interested in probability densities since the "probability density constraint" will involve the constraint function $f_1 \equiv 1$ which is not in L^2, except in the special case where we impose the density to have a compact support. To overcome that difficulty, we consider in the second subsection the same optimization problem with constraint functions f_i that do not necessarily belong to L^2. We show that the projection on S has the same form as before as soon as there exist coefficients α_i^* such that $\int (f - \sum_{i=1}^n \alpha_i^* f_i)^+ f_i = d_i$ for $i = 1, \cdots, n$. The last subsection considers the special case of two contraints with contraints functions $f_1 \equiv 1\,,\ f_2 \geq 0$ not necessarily in L^2 and gives the conditions on (d_1, d_2) under which there exists coefficients α_i^* such that $\int (f - \sum_{i=1}^n \alpha_i^* f_i)^+ f_i = d_i$ for $i = 1, 2$ and therefore under which the solution can be expressed as $g^* = (f - \alpha_1^* f_1 - \alpha_2^* f_2)^+$. Hence, the functional optimization problem reduces to determine the real valued coefficients α_i^*.

3.1 The case of L^2 constraints

The projection problem (3) can be viewed as an optimization problem with several linear constraints and a positivity contraint. The following proposition gives the expression of the projection g^* of f on S using Lagrange approach.

Proposition 3.1. *Let $f \in L^2(\mathbb{R}^d, \mathbb{R})$, consider the following optimization problem:*

$$\min_{g \geq 0} \int_{\mathbb{R}} |f-g|^2 \ \text{ with } \int g f_i = d_i \ \text{ for } \ i = 1, ..., n \ , \tag{4}$$

where $(f_1, \cdots, f_n)$ are linearly independent functions in $L^2(\mathbb{R}^d, \mathbb{R})$, then there exists $(\alpha_1^*, \cdots, \alpha_n^*) \in \mathbb{R}^n$ such that the solution of (4)

$$g^* = (f - \sum_{i=1}^n \alpha_i^* f_i)^+ \ . \tag{5}$$

PROOF The Lagrangian of (4) is :

$$Ł = \frac{1}{2} \|f-g\|^2 - \sum_{i=1}^n \lambda_i (d_i - \langle f_i, g \rangle) - \mu g \ , \tag{6}$$

where $\langle f_i, g \rangle = \int f_i g$ is the usual scalar product in L^2 and μ a function on $\mathbb{R}^d$. With the assumption that the functions f_i belong to L^2, the theory states [5] that (4) has a unique solution g^* given by the necessarily and sufficient following conditions :

$$\begin{cases} g^* - f - \sum_{i=1}^n \lambda_i f_i - \mu = 0 \\ \mu g^* = 0 \\ \mu \geq 0 \\ g^* \geq 0 \ . \end{cases} \tag{7}$$

where μ is a function. Let us introduce the set $A = \{x \,|\, g^*(x) > 0\}$. The second equation of (7) implies $\mu = 0$ on A. Then, using the first equation of (7) implies the following equality on the subspace A

$$g^* \mathbf{1}_A = (f + \sum_{i=1}^n \lambda_i f_i) \mathbf{1}_A = (f + \sum_{i=1}^n \lambda_i f_i)^+ \mathbf{1}_A \ .$$

Since $g^* \geq 0$ on $\mathbb{R}^d$ then $g^* \equiv 0$ outside of A. On the other hand, the first equation of (7) yields that on the complement of A, denoted $\bar{A}$ the following inequality holds

$$(f + \sum_{i=1}^n \lambda_i f_i) \mathbf{1}_{\bar{A}} = -\mu \mathbf{1}_{\bar{A}} \leq 0 \ .$$

Hence

$$g^* \mathbf{1}_{\bar{A}} = 0 = (f + \sum_{i=1}^n \lambda_i f_i)^+ \mathbf{1}_{\bar{A}} \ .$$

This finally yields the following equality valid on the whole space $\mathbb{R}^d$,

$$g^* = (f + \sum_{i=1}^n \lambda_i f_i)^+ \ .$$

Then we end the proof by setting $\alpha_i = -\lambda_i$ for $i = 1, \cdots, n$.

3.2 The general case

Proposition 3.1 gives the form of the solution of the problem (3) when the constraint functions $f_1, \cdots, f_n$ are in L^2 without any other conditions. To be able to take into account the "probability density constraint" corresponding to $f_1 \equiv 1$ and $d_1 = 1$ we would like to relax the assumption that the constraint functions $f_1, \cdots, f_n$ belong to L^2. The following proposition achieves this goal.

Proposition 3.2. *Let f be a function defined on $\mathbb{R}^d$ with values in $\mathbb{R}$, such that $\int_{\mathbb{R}^d} |f(x)|^2 \, dx < \infty$, and S be the closed convex set defined in (2). Assume that there exists reals $(\alpha_1^*, \cdots, \alpha_n^*)$ such that for $i = 1, \cdots, n$*

$$\int (f - \sum_{i=1}^n \alpha_i^* f_i)^+ f_i = d_i \ . \tag{8}$$

Then the projection g^* of f on the subspace S of L^2-probability densities satisfying the linear constraints (2), is determined by

$$g^* = (f - \sum_{i=1}^n \alpha_i^* f_i)^+ = \arg\min_{g \in S} \int |f-g|^2 \ . \tag{9}$$

PROOF The proof is based on the characterization of the projection X^* of a point Y on a closed convex set S in a Hilbert space [9]. X^* is the projection of Y on S if and only if

$$\langle Y - X^*, X - X^* \rangle \leq 0 \quad \text{for all} \quad X \in S .$$

In our case, S is the subspace of L^2-probability densities with linear constraints (2). It is sufficient to prove that,

$$\Delta = \int [f - (f - \sum_{i=1}^{n} \alpha_i^* f_i)^+] [g - (f - \sum_{i=1}^{n} \alpha_i^* f_i)^+] \qquad (10)$$

is negative for any probability density $g \in S$, where $\alpha_1^*, \cdots, \alpha_n^*$ are defined in (8). Let us introduce the subsets of $\mathbb{R}^d$,

$$A^* = \{x \,|\, f(x) \geq \sum_{i=1}^{n} \alpha_i^* f_i(x)\} \text{ and } \bar{A}^* = \{x \,|\, f(x) < \sum_{i=1}^{n} \alpha_i^* f_i(x)\} .$$

Splitting expression (10) of Δ into two terms yields

$$\begin{aligned}
\Delta &= \int_{A^*} [f - f + \sum_{i=1}^{n} \alpha_i^* f_i] [g - f + \sum_{i=1}^{n} \alpha_i^* f_i] + \int_{\bar{A}^*} fg \\
&= \sum_{i=1}^{n} \alpha_i^* \left[\int_{A^*} g f_i - \int (f - \sum_{i=1}^{n} \alpha_i^* f_i)^+ f_i \right] + \int_{\bar{A}^*} fg \\
&= \sum_{i=1}^{n} \alpha_i^* \left[\int_{A^*} g f_i - d_i \right] + \int_{\bar{A}^*} fg \\
&= -\sum_{i=1}^{n} \alpha_i^* \int_{\bar{A}^*} g + \int_{\bar{A}^*} fg = \int_{\bar{A}^*} g (f - \sum_{i=1}^{n} \alpha_i^* f_i)
\end{aligned}$$

This quantity is non-positive since $g \in S$ is a positive function, which ends the proof.

3.3 Existence of positive part like solution with two linear constraints

In general, it can be difficult to check whether the existence condition (8) is satisfied or not. The following proposition gives a necessary and sufficient condition under which condition (8) is satisfied in the case of two linear constraints : one constraint corresponding to the density constraint ($f_1 \equiv 1$, $d_1 = 1$) and one general constraint with a non-negative function f_2. For simplicity, thereafter the general notation f_1 is kept, although $f_1 \equiv 1$. Moreover, only the case where α_i^* are positive is considered.

Proposition 3.3. *Let f be a bounded function in $L^2(\mathbb{R}^d)$, consider the following constrained optimization problem:*

$$\min_{g \geq 0} \int |f - g|^2 \;\; \textit{with} \;\; \int g f_i = d_i \;\; \textit{for} \;\; i = 1, 2 , \qquad (11)$$

with f_2 a non-negative measurable function not necessarily lying in L^2 and $f_1 \equiv 1$. Assume that the following condition is verified:

$$\begin{cases} 0 \leq d_1 \leq \int f^+ f_1 < \infty , \\ 0 \leq d_2 \leq \int f^+ f_2 < \infty . \end{cases} \qquad (12)$$

Then the solution of (11) has the following form

$$g^* = (f - \alpha_1^* f_1 - \alpha_2^* f_2)^+ , \qquad (13)$$

where α_i^* are positive scalars (not necessarily unique) such that

$$\int g^* f_i = d_i \quad \textit{for} \;\; i = 1, 2 , \qquad (14)$$

if and only if (d_1, d_2) verify one of the following conditions:

- *Case 1 : $\int_{f_2=0} f^+ f_1 \leq d_1$. The condition on d_2 is,*

$$\int (f - \bar{\alpha}_2 f_2)^+ f_2 \leq d_2 \leq \int (f - \bar{\alpha}_1 f_1)^+ f_2 . \qquad (15)$$

  where $(\bar{\alpha}_1, \bar{\alpha}_2)$ are the non-negative solutions of

$$\begin{cases} \int (f - \bar{\alpha}_1 f_1)^+ f_1 = d_1 \\ \int (f - \bar{\alpha}_2 f_2)^+ f_1 = d_1 . \end{cases} \qquad (16)$$

  In that case $\alpha_1^* \in [0, \bar{\alpha}_1]$ and $\alpha_2^* \in [0, \bar{\alpha}_2]$.

- *Case 2: $\int_{f_2=0} f^+ f_1 > d_1$. The condition on d_2 is,*

$$0 \leq d_2 \leq \int (f - \bar{\alpha}_1 f_1)^+ f_2 . \qquad (17)$$

  where $\bar{\alpha}_1$ and $\bar{\bar{\alpha}}_1$ are the non-negative solution of

$$\begin{cases} \int (f - \bar{\alpha}_1 f_1)^+ f_1 = d_1 \\ \int_{f_2=0} (f - \bar{\bar{\alpha}}_1)^+ = d_1 . \end{cases} \qquad (18)$$

  In that case $\alpha_1^* \in [\bar{\bar{\alpha}}_1, \bar{\alpha}_1]$ and $\alpha_2^* \in [0, \infty)$.

PROOF By Proposition 3.2, it is sufficient to proof the existence of α_1^* and α_2^* satisfying (14). Thereafter we use the following technical result (the proof is left to the reader): let $f \in L^2(\mathbb{R}^d)$ a bounded function with f_2 a non negative function and $\alpha_2 \geq 0$ then, $k(\alpha_2) = \int (f - \alpha_2 f_2)^+$ is a non-increasing continuous function varying from $k(0) = \int f^+$ to $k(+\infty) = \int_{x/f_2(x)=0} f^+$.

Let us introduce the following non-negative function, defined on $\mathbb{R}^+ \times \mathbb{R}^+$,

$$H(\alpha_1, \alpha_2) = \int (f - \alpha_1 f_1 - \alpha_2 f_2)^+ f_1 . \qquad (19)$$

f_1 and f_2 being non-negative functions, the maximum value of $H(\alpha_1, \alpha_2)$ on $\mathbb{R}^+ \times \mathbb{R}^+$ is $\int f^+ f_1$ and is obtained for $(\alpha_1, \alpha_2) = (0, 0)$. Now, let us consider the following equation,

$$H(\alpha_1, \alpha_2) = d_1 . \qquad (20)$$

The function $\alpha_2 \mapsto H(0, \alpha_2) = \int (f - \alpha_2 f_2)^+ f_1$ decreases continuously as α_2 increases, from $H(0,0) = \int f^+ f_1$ to $H(0, \infty) = \int_{f_2=0} f^+ f_1$. Now, let us consider the two following cases,

- Case 1 : $\int\limits_{f_2=0} f^+ f_1 \leq d_1$. From (12) we deduce that $H(0,0) \geq d_1$ while $H(0,\infty) \leq d_1$. Therefore it exists $\bar{\alpha}_2$ such that $H(0,\bar{\alpha}_2) = d_1$. For a given $\alpha_2 \in \mathbb{R}^+$, there is a solution $\alpha_1 = \alpha_1(\alpha_2)$ of equation (20) iff $\alpha_2 \leq \bar{\alpha}_2$. Indeed, if $\alpha_2 \leq \bar{\alpha}_2$, $\alpha_1 \mapsto H(\alpha_1,\alpha_2)$ decreases from $H(0,\alpha_2) \geq d_1$ to $H(\infty,\alpha_2) = 0$ (recall that $f_1 \equiv 1$ and that f is bounded). Hence, for any $\alpha_2 \in [0,\bar{\alpha}_2]$, there exists $\alpha_1(\alpha_2) \in [0,\bar{\alpha}_1]$ such that

$$H(\alpha_1(\alpha_2),\alpha_2) = d_1 , \tag{21}$$

where $\bar{\alpha}_1$ and $\bar{\alpha}_2$ are given by equation (16). Notice that in that case $\alpha_1^* \in [0,\bar{\alpha}_1]$.
- Case 2 : $\int\limits_{f_2=0} f^+ f_1 > d_1$. From (12), we deduce that $H(0,0) \geq d_1$ while $H(0,\infty) > d_1$. Hence, $H(0,\alpha_2) \geq d_1$ for any $\alpha_2 \geq 0$. $(\alpha_2,\, \alpha_1) \mapsto H(\alpha_1,\alpha_2)$ decreases from $H(0,\alpha_2) > d_1$ to $H(\infty,\alpha_2) = 0$. Hence, for any $\alpha_2 \in \mathbb{R}^+$, there exists $\alpha_1(\alpha_2) \in \mathbb{R}^+$ such that

$$H(\alpha_1(\alpha_2),\alpha_2) = d_1 . \tag{22}$$

Notice that $\alpha_2 \mapsto \alpha_1(\alpha_2)$ is decreasing from $\alpha_1(0) = \bar{\alpha}_1$ to $\alpha_1(\alpha_2 = \infty) = \bar{\bar{\alpha}}_1$ where $\bar{\bar{\alpha}}_1$ is solution of $\int\limits_{f_2=0} (f - \bar{\bar{\alpha}}_1)^+ f_1 = d_1$. Hence in that case $\alpha_1^* \in [\bar{\bar{\alpha}}_1, \bar{\alpha}_1]$.

Now, let us deal with the second constraint. Let us introduce the following non-negative function with $0 \leq \alpha_2 \leq \bar{\alpha}_2$ for the case 1 and $0 \leq \alpha_2 \leq \infty$ for the case 2:

$$G(\alpha_2) = \int (f - \alpha_1(\alpha_2) f_1 - \alpha_2 f_2)^+ f_2 . \tag{23}$$

- Case 1 : G varies continuously from $G(0) = \int (f - \bar{\alpha}_1 f_1)^+ f_2$ to $G(\bar{\alpha}_2) = \int (f - \bar{\alpha}_2 f_2)^+ f_2$. That is, $G(\alpha_2) = d_2$ has a solution iff $G(\bar{\alpha}_2) \leq d_2 \leq G(0)$, which gives the condition (15) on d_2.
- Case 2 : G varies continuously from $G(0) = \int (f - \bar{\alpha}_1 f_1)^+ f_2$ to $G(\alpha_2 = \infty) = 0$. Then notice that $\lim\limits_{\alpha_2 \to +\infty} \int\limits_{f_2>0} (f - \alpha_1(\alpha_2) f_1 - \alpha_2 f_2)^+ f_2 = 0$ observing that $\lim\limits_{\alpha_2 \to +\infty} \alpha_1(\alpha_2) = \bar{\bar{\alpha}}_1 < \infty$. We can conclude that $G(\alpha_2) = d_2$ has a solution iff $d_2 \leq G(0)$, which gives the condition (17) on d_2.

Observing that $G(\alpha_2)$ is continuous, we conclude that $G(\alpha_2) = d_2$ has at least one solution $\alpha_2 = \alpha_2^*$ under the conditions (15) - (17). Then, $\alpha_1^* = \alpha_1(\alpha_2^*)$ is the associated solution of (22). Applying proposition 3.2, we conclude the proof of proposition : the solution of the constrained optimization problem (11) has the form (13) with α_i^* positive iff (d_1, d_2) satisfy the conditions (12) - (18).
Moreover, one can prove that the solution $(\alpha_1^*,\ \alpha_2^*)$ is unique iff f_1 and f_2 are linearly independent on $\Omega(\alpha) = \{x \,|\, f(x) - \alpha_1 f_1(x) - \alpha_2 f_2(x) \geq 0\}$. Thereafter, we present a sketch of the proof by demonstrating that $G(\alpha_2)$ is a strictly decreasing function. First, applying the implicit function theorem to $H(\alpha_1(\alpha_2),\alpha_2) = d_1$ gives :

$$\frac{d\alpha_1(\alpha_2)}{d\alpha_2} = -\left(\frac{\partial H}{\partial \alpha_1}\right)^{-1} \left.\frac{\partial H}{\partial \alpha_2}\right|_{\alpha_1 = cte} \tag{24}$$

By developing $G(\alpha_2 + \varepsilon)$ around α_2 for small ε, we obtain,

$$\frac{dG(\alpha_2)}{d\alpha_2} = -\int_{\Omega(\alpha)} (\frac{d\alpha_1}{d\alpha_2} f_1 + f_2) f_2 \tag{25}$$

$$\frac{\partial H}{\partial \alpha_1} = -\int_{\Omega(\alpha)} f_1^2 \quad \text{and} \quad \frac{\partial H}{\partial \alpha_2} = -\int_{\Omega(\alpha)} f_1 f_2 .$$

Injecting (24) in (25) we finally express the gradient of G,

$$\frac{\partial G}{\partial \alpha_2} = \frac{\left(\int_{\Omega(\alpha)} f_1 f_2\right)^2 - \int_{\Omega(\alpha)} f_1^2 \int_{\Omega(\alpha)} f_2^2}{\int_{\Omega(\alpha)} f_1^2} . \tag{26}$$

Thanks to Schwarz inequality we see that, for all α_2, $\frac{\partial G}{\partial \alpha_2} < 0$ as soon as f_1 and f_2 are linearly independent on $\Omega(\alpha)$.

4. SIMULATIONS

4.1 Application to support constraint

Let f an $L^2(\mathbb{R})$ function with any support (possibly infinite). The goal is to approximate f by a density having a compact support C. This projection method can be applied when we know a priori the support of the true density. This is the case for example with salary data which are necessarily positive [8]. For this application, if we approximate this density by a kernel estimation density, the result can provide negative values. Now, we set one constraint $f_1 \equiv \mathbf{1}$ (which does not belong to $L^2(\mathbb{R})$) with $d_1 = 1$ for the "density constraint" and $f_2 \equiv \mathbf{1}_{x \in C}$ with $d_2 = 1$ for the support constraint. We focus on the case 1 of the proposition 3.3 which give, with the necessary conditions (12) : $\int_{\bar{C}} f^+ \leq 1 \leq \int_C f^+$ where $\bar{C}$ is the complement of C. The conditions (15) for the existence of α_i^* give:

$$\int_C (f - \bar{\alpha}_2)^+ \leq 1 \leq \int_C (f - \bar{\alpha}_1)^+ . \tag{27}$$

where $(\bar{\alpha}_1, \bar{\alpha}_2)$ verify (16):

$$\begin{cases} \displaystyle\int_C (f - \bar{\alpha}_1)^+ + \int_{\bar{C}} (f - \bar{\alpha}_1)^+ = 1 , \\ \displaystyle\int_C (f - \bar{\alpha}_2)^+ + \int_{\bar{C}} f^+ = 1 \end{cases} \tag{28}$$

(27) and (28) imply $\int_{\bar{C}} (f - \bar{\alpha}_1)^+ = 0$ and $\int_C (f - \bar{\alpha}_1)^+ = 1$ which gives $\bar{\alpha}_1 \geq \sup\limits_{x \in \bar{C}} f(x)$. This implies the condition,

$$\int_C (f - \sup_{x \in \bar{C}} f(x))^+ \geq 1, \tag{29}$$

Under this condition, we set $\alpha_1^* = \sup\limits_{\bar{C}} f$ and α_2^* such that $\int_C (f - \sup\limits_{x \in \bar{C}} f(x) - \alpha_2^*)^+ = 1$, we obtain the expression of g^*:

$$g^*(x) = (f(x) - \sup_{x \in \bar{C}} f(x) - \alpha_2^* \mathbf{1}_{x \in C})^+ \tag{30}$$

We can check that g^* is a density which is zero in $\bar{C}$.

As an illustrative example we take $f(x) = \dfrac{1}{\pi}\dfrac{\sin(x)}{x}$ centered in $x = 10$. We wish to project f on the densities space with support $C = [0, 25]$. This function satisfies the condition (29). On figure 1, g^* is compared with the so called L^1-projection $g_1^*(x) = \mathbf{1}_{x \in C} f^+(x) / \int_C f^+(x)$. Simulation results show that the L^1 error is identical for the 2 projections, which agrees with [10], but the L^2 error is 0.18 for g_1^* and 0.15 for g^*.

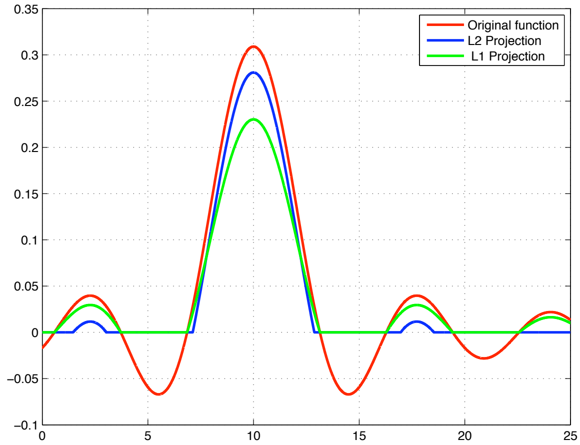

Figure 1: Original function compared with the 2 projections. (Case constraint functions not belonging to L^2)

4.2 Application to mean constraint

Here, we will consider a function f with compact support C. Let us consider again a case where $f_1 \equiv 1$ and $d_1 = 1$ for the density constraint and $f_2 \equiv x$ and $d_1 = \mu$ for the mean constraint. Here the constraint functions are in $L^2(C)$ and the Lagrangian method gives (Proposition 3.1):

$$g^*(x) = (f(x) - \alpha_1 - \alpha_2 x))^+ \qquad (31)$$

where α_i (possibly negative) are determined by the 2 constraints. As an illustrative example we take $f(x) = k(\varphi_1(x) + \varphi_2(x))$ where $\varphi_i = N(\beta_i, \sigma_i^2)$ are 2 Gaussian densities of mean and standard deviation $(\beta_1 = 1, \sigma_1 = 1)$ and $(\beta_2 = 3, \sigma_2 = 0.2)$. The support of f is $C = [\beta_1 - 3\sigma_1, \beta_1 + 3\sigma_1]$ and k is such that f is a density. The mean of f is 2 and the imposed mean is $\mu = 3$. On figure 2, g^* is compared with a test function of the form $g_1 = \gamma_1 f(x - \gamma_2)$ where γ_i are such that g_1 satisfies the 2 constraints. Simulation results show that the L^1 and L^2 errors are drastically reduced. The L^1 error is 0.7 for g^*, 1 for g_1 and the L^2 error is 0.35 for g^*, 0.55 for g_1.

5. CONCLUSION

This paper presents a quite general methodology to estimate a density under linear constraints (for example: positivity, support, moments or quantiles). The desired approximation is the projection of the free density estimation on the set of the probability densities satisfying the constraints. The solution is expressed in a simple form. Some simulations show the improvement of the approximation compared with alternative methods.

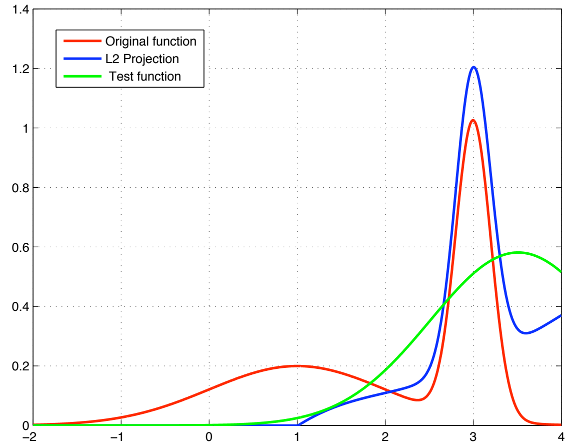

Figure 2: Original function compared with the projection and the test function. (Case constraint functions belonging to L^2)

Acknowledgments

The authors would like to thank Kengy Barty and Arnaud Lenoir from EDF-R&D for useful discussions on Lagrangian methods. The authors thank also Bobby Rangell to have proof-read this paper.

REFERENCES

[1] L. Devroye, L. Gyorfi. *Nonparametric Density Estimation : The L_1 View*. John Wiley & Sons.Wiley series in Probability and Mathematical Statistics, New York 1985.

[2] B.W. Silverman. *Density Estimation for Statistics and Data Analysis*. Chapman & Monographs on Statistics and Applied Probability, London 1986.

[3] L. Devroye. *A Course on Density Estimation*.Birkhuser, Boston 1987.

[4] R.J. Serfling . *Asymptotic expansion*. Encyclopedia of Statistical Sciences 1 137-138, J. Wiley&Sons, New York 1982.

[5] V. Alexeev, E. Galeev and V. Tikhomorov . *Recueil de problemes d'optimisation*. Editions Mir, Moscou 1987

[6] E. Parzen, "On estimation of a probability density and mode" *Annals of Mathematical Statistics*, vol. 33, pp. 1065-1070, 1962.

[7] M.S. Bartlett , "Statistical estimation of density functions" *Sankhya Ser. A, vol*. 2, pp. 1245-254, 1963.

[8] P. Hall , B. Presnell, "Density Estimation Under Constraints" *Journal of Computational and Graphical Statistics*, vol. 8, Number 2, pp. 259-277, 1999.

[9] C.K. Chui, F.D. Deutsch and J.D. Ward, "Constrained Best Approximation in Hilbert Space" *Constructive Approximation*, vol. 6, pp. 35-64, 1990.

[10] N. Oudjane, C. Musso, "L^2-Density estimation with negative kernels," in *Proc. ISPA 2005*, Zagreb, Croatia, September 15-17, 2005.