



**HAL**  
open science

## Prediction based on a L1-method in the nonlinear autoregressive model

Benoît Cadre, Irène Larramendy

► **To cite this version:**

Benoît Cadre, Irène Larramendy. Prediction based on a L1-method in the nonlinear autoregressive model. *Mathematical Methods of Statistics*, 2006, pp.256-268. hal-00455752

**HAL Id: hal-00455752**

**<https://hal.science/hal-00455752v1>**

Submitted on 11 Feb 2010

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

PREDICTION BASED ON A  $L_1$ -METHOD IN THE  
NONLINEAR AUTOREGRESSIVE MODEL

**Running head** : Forecasting by a  $L_1$ -method

Benoît Cadre, Irène Larramendy-Valverde  
Laboratoire de Probabilités et Statistiques  
CC 051, Place Eugène Bataillon  
34095 MONTPELLIER Cedex 05  
FRANCE

**Abstract** We consider in this paper the  $d$ -dimensional nonlinear autoregressive model of order 1 defined by:

$$X_{n+1} = F(X_n) + \sigma(X_n)\varepsilon_{n+1}, \quad n \geq 0,$$

where  $(\varepsilon_n)_{n \geq 1}$  is a sequence of independent, identically distributed  $\mathbb{R}^d$ -valued random variables, and  $F : \mathbb{R}^d \rightarrow \mathbb{R}^d$  (resp.  $\sigma : \mathbb{R}^d \rightarrow \mathbb{R}_+$ ) is a Hölder function. From a  $L_1$ -method, we construct a robust and uniformly consistent estimator  $m_n$  of the function  $F + \sigma\alpha_0$ , where  $\alpha_0 \in \mathbb{R}^d$  is the unique  $L_1$ -median of  $\varepsilon_1$ . Since  $F(X_n) + \sigma(X_n)\alpha_0$  is the best predictor (for the  $L_1$ -norm) of  $X_{n+1}$  we can get when  $X_n$  is known, the estimator  $m_n$  provides the statistical predictor  $m_n(X_n)$ . Numerical simulations are provided, inciting the choice of  $L_1$ -method for forecasting.

**AMS 2000 subject classification** : 62M10, 62M20.

**Key-words and phrases** : nonlinear model,  $L_1$ -median, empirical estimator, stationary distribution, statistical prediction.

## Introduction

We consider the  $d$ -dimensional nonlinear autoregressive model of order 1  $(X_n)_{n \geq 0}$  defined by:

$$X_{n+1} = F(X_n) + \sigma(X_n)\varepsilon_{n+1}, \quad n \geq 0,$$

where  $F : \mathbb{R}^d \rightarrow \mathbb{R}^d$ ,  $\sigma : \mathbb{R}^d \rightarrow \mathbb{R}_+$ ,  $(\varepsilon_n)_{n \geq 1}$  is a sequence of independent, identically distributed  $\mathbb{R}^d$ -valued random variables such that for all  $n \geq 0$ ,  $\varepsilon_{n+1}$  is independent of  $X_0, \dots, X_n$ .

The aim of the paper is to construct a statistical predictor of  $X_{n+1}$  from the data  $X_0, \dots, X_n$ . Mostly, the predictors of  $X_{n+1}$  are defined by a  $L_2$ -method (i.e. using a quadratic minimizing criterion) and it is often assumed that  $(X_n)_{n \geq 0}$  is a strictly stationary and mixing sequence (see for instance Brockwell et al [5] for the linear case and Diebolt [8], Diebolt et al [9], Doukhan et al [11], Robinson [17] for the nonlinear and univariate case). On one hand, the  $L_2$ -methods do not give robust estimators and a predictor based on a  $L_1$ -method should give a better prediction, just because the  $L_1$ -norm is smaller than the  $L_2$ -norm. On the other hand, the stationarity property on the sequence  $(X_n)_{n \geq 0}$  turns to be a very restrictive assumption (for instance, one has to assume that the law of  $X_0$  is the law of the unknown stationarity distribution, when it exists). Moreover, the mixing condition on the sequence  $(X_n)_{n \geq 0}$  requires absolute continuity of  $\varepsilon_1$ 's density and the particular form of the mixing coefficients involves additional assumptions on the model (see Doukhan [10]).

In this article, we eliminate these severe drawbacks. We introduce another predictor based on the  $L_1$ -median. The  $L_1$ -median provides a robust estimator (see Kemperman ([15], Theorem 3.10)) and the statistical predictor defined by this  $L_1$ -method should give a better predictor than the one given by the  $L_2$ -method, at least because the  $L_1$ -norm is smaller than the  $L_2$ -norm. Moreover, we shall not assume in this paper that the sequence  $(X_n)_{n \geq 0}$  is strictly stationary and mixing which, as mentioned above, involves very restrictive assumptions on the model. However, we shall only prove the consistency of our estimator, which is much more difficult than in the above cases, just because the predictor may not be defined via a simple expression. The numerical study of Section 2 proves the very good performances of our predictor, compared to its analogous predictor constructed via a  $L_2$ -method.

Throughout the paper, we let  $\|\cdot\|$  to be a fix norm on  $\mathbb{R}^d$ . Recall that the set of  $L_1$ -medians of an integrable bounded measure  $\nu$  defined on  $\mathbb{R}^d$  is:

$$\text{Argmin}_{\alpha \in \mathbb{R}^d} \int \|x - \alpha\| \nu(dx).$$

One easily prove that the set of  $L_1$ -medians is always non-empty. For statistical uses of the  $L_1$ -medians, we refer to the survey by Small [18] (see also Berline et al [1] and [2], Gannoun et al [13]).

Let us now explain the basic idea of the paper. Assume that the law of  $\varepsilon_1$  has a unique  $L_1$ -median  $\alpha_0 \in \mathbb{R}^d$  (by Kemperman ([15], Theorem 3.17) or Milasevic and Ducharme [16], this is the case if for instance,  $\|\cdot\|$  is strictly convex - i.e.  $\|x + y\| < \|x\| + \|y\|$  whenever  $x$  and  $y$  are not proportional - and if the support of the law of  $\varepsilon_1$  is not included into a straight line). Then, the law of  $X_1$  given  $X_0 = x$  has a unique  $L_1$ -median which is easily seen to be  $F(x) + \sigma(x)\alpha_0$ . One can now introduce an estimate of  $F + \sigma\alpha_0$  from the data  $X_0, \dots, X_n$ . We first consider a Nadaraya-Watson type estimator of the law of  $X_1$  given  $X_0 = x$ , namely  $\nu_n^x$ . We then prove that with probability 1 and uniformly on  $x$ , the set of  $L_1$ -medians of  $\nu_n^x$ , called  $m_n(x)$ , converges to  $F(x) + \sigma(x)\alpha_0$ . Since  $F(X_n) + \sigma(X_n)\alpha_0$  is the best predictor (for the  $L_1$ -norm) of  $X_{n+1}$  we can get when  $X_n$  is known, the estimator  $m_n$  provides the statistical predictor  $m_n(X_n)$ .

The paper is organized as follows. Notations, hypotheses and main results are given in Section 1. Some numerical simulation are presented in Section 2. Section 3 is devoted to the proof of the main result.

## 1. Notations, hypotheses and main results

### 1.1 Notations and hypotheses

*Notations and basic assumptions on the model.* From now on,  $\varepsilon$  is a  $\mathbb{R}^d$ -valued random variable with the same law as  $\varepsilon_1$ . We assume throughout that  $\varepsilon$  has a positive, Lipschitz density  $g$ , and  $E[\|\varepsilon\|^3] < \infty$ . Moreover, we assume that  $\sigma$  and  $F$  are Hölder functions of order  $H_1 > 0$  and  $H_2 > 0$ ,  $\inf \sigma > 0$  and we make the assumption that the model is stable i.e. there exists a probability  $\mu$  on  $\mathbb{R}^d$  such that with probability 1, the sequence of probability measures

$$\frac{1}{n} \sum_{i=0}^n \delta_{X_i}, \quad n \geq 1$$

weakly converges to  $\mu$ , where  $\delta_z$  denotes the Dirac measure on  $z \in \mathbb{R}^d$ . We finally assume that the law of  $\varepsilon$  has only one  $L_1$ -median (for a sufficient criterion of uniqueness, see Kemperman [15]). We let  $\alpha_0 \in \mathbb{R}^d$  be the  $L_1$ -median of  $\varepsilon$  and for all  $x \in \mathbb{R}^d$ , we let  $\nu^x$  to be the law of  $X_1$  given  $X_0 = x$ . Note that since  $X_0$  is independent of  $\varepsilon_1$ ,  $\nu^x$  is also the law of  $F(x) + \sigma(x)\varepsilon$ . If moreover  $\alpha \in \mathbb{R}^d$ , we let:

$$\varphi(\alpha|x) = \int \|y - \alpha\| \nu^x(dy) = E[\|F(x) + \sigma(x)\varepsilon - \alpha\|].$$

**Remark 1.1** If  $\sup \sigma < \infty$  and there exists  $L < 1$  such that

$$\|F(x)\| \leq L\|x\|^{H_1}, \quad \forall x \in \mathbb{R}^d,$$

then, we deduce from an easy modification of Duflo ([12], p. 192, Example 2 - in which the case  $H_1 = 1$  is considered) that the model is stable.

*The empirical estimators.* With the convention  $0/0 = 0$ , the estimator of  $\nu^x$  is defined for all  $n \geq 1$  by:

$$\nu_n^x = \frac{\sum_{i=1}^n \delta_{X_i} i^{ad} K(i^a \|X_{i-1} - x\|)}{\sum_{i=1}^n i^{ad} K(i^a \|X_{i-1} - x\|)},$$

where  $a > 0$  and  $K : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  is a Lipschitz density function with a compact support contained in  $[0, 1]$  (for simplicity of the proofs). Moreover, for  $\alpha, x \in \mathbb{R}^d$  and  $n \geq 1$ , we let:

$$\varphi_n(\alpha|x) = \int \|y - \alpha\| \nu_n^x(dy).$$

With probability 1, for all  $x \in \mathbb{R}^d$  and  $n \geq 1$ , the set of  $L_1$ -medians of  $\nu_n^x$ , i.e.:

$$\text{Argmin}_{\alpha \in \mathbb{R}^d} \varphi_n(\alpha|x) = \text{Argmin}_{\alpha \in \mathbb{R}^d} \sum_{i=1}^n i^{ad} \|X_i - \alpha\| K(i^a \|X_{i-1} - x\|),$$

is non-empty. We let  $m_n(x)$  to be one of the  $L_1$ -medians of  $\nu_n^x$ .

## 1.2 The robust estimator

**Theorem 1.1** *Assume that  $\sup_{n \geq 1} E[\|X_n\|] < \infty$ . If  $ad(4+d) < 1$  then, with probability 1 (w.p. 1):*

$$\|m_n - (F + \sigma \alpha_0)\| \rightarrow 0, \quad \text{as } n \rightarrow \infty,$$

*uniformly on every compact set.*

**Remark 1.2** The condition on the sequence  $(E[\|X_n\|])_{n \geq 1}$  always holds in the following important cases :

- $X_0$  is integrable and its law is given by  $\mu$  (because  $(X_n)_{n \geq 0}$  is then a strictly stationary sequence);
- $F$  and  $\sigma$  are bounded;
- $F$  and  $\sigma$  satisfy the conditions of Remark 1.1 (see Duflo ([12], p. 192)).

**Remark 1.3** Several improvements of Theorem 1.1 may be considered. For instance, one may consider the Banach AR(1) model (see Bosq [3]), or even a functional autoregressive model defined on a Banach space. However, the convergence is then obtained only for a weak topology (see Cadre [6] for a result in this direction).

**Remark 1.4** If we assume moreover that  $d \geq 2$  and the sequence  $(X_n)_{n \geq 0}$  is strictly stationary and geometrically mixing then, one can prove using the Hölder property of the  $L_1$ -median given in Cadre [7] (for similar computations, see also the rate of convergence for the conditional  $L_1$ -median in the same paper) that w.p. 1:

$$\|m_n - (F + \sigma\alpha_0)\| = O(n^{-(1-ad)/4}),$$

uniformly on every compact set. In this particular framework, this gives a non-optimal rate of convergence. The proof of this result is very different from the proof of Theorem 1.1 : here, we insist again on the fact that Theorem 1.1 is obtained under mild assumptions on the model.

### 1.3 Statistical prediction

One can easily prove that, for the  $L_1$ -norm,  $F(X_n) + \sigma(X_n)\alpha_0$  is the best approximation of  $X_{n+1}$  we can get when  $X_n$  is known, i.e.:

$$E[\|X_{n+1} - (F(X_n) + \sigma(X_n)\alpha_0)\| | X_n] = \inf_{\alpha \in \mathbb{R}^d} E[\|X_{n+1} - \alpha\| | X_n].$$

Hence, the estimator  $m_n$  of  $F + \sigma\alpha_0$  provides the statistical predictor  $m_n(X_n)$ . Note that if  $\sup_{n \geq 1} E[\|X_n\|] < \infty$ , then the sequence  $(X_n)_{n \geq 1}$  is tight. The corollary below is then straightforward from Theorem 1.1.

**Corollary 1.1** *Assume that  $\sup_{n \geq 1} E[\|X_n\|] < \infty$ . If  $ad(4+d) < 1$  then:*

$$m_n(X_n) - (F(X_n) + \sigma(X_n)\alpha_0) \rightarrow 0, \text{ as } n \rightarrow \infty,$$

*in probability.*

In Section 2 below, we shall compare the performances of our predictor to the performances of the analogous predictor constructed via a  $L_2$ -method (see for example Robinson [17]).

## 2. Simulation study and data analysis

### 2.1 Notations and mean error of prediction

Let  $n$  be the sample size and  $x_1, \dots, x_n$  the observations. We define  $\hat{x}_j^1$  (respectively  $\hat{x}_j^2$ ) the prediction of  $x_j$  constructed from  $x_1, \dots, x_{j-1}$  with the classical kernel predictor (see for example Robinson [17]) (respectively with the predictor defined in section 1.3.). When the number  $H$  of predictions is given, we then compute the mean error of the predictions defined by

$$EK = \frac{1}{H} \sum_{j=n-H+1}^n \frac{\|x_j - \hat{x}_j^1\|}{\|x_j\|}, \quad EM = \frac{1}{H} \sum_{j=n-H+1}^n \frac{\|x_j - \hat{x}_j^2\|}{\|x_j\|}.$$

In the numerical study, we choose the gaussian kernel  $K$  defined by

$$K(t) = \frac{1}{(2\pi)^{d/2}} \exp\left(-\frac{1}{2}\|t\|^2\right).$$

### 2.2 Simulations

In the univariate case ( $d = 1$ ), a simulation study is undertaken to investigate the methodology presented in the previous section. The different link functions  $F$  considered are :

$$F(x) = ax + 10 \text{ with } a \in \{0.8; 0.9; 1; 1.02\} \text{ (model A);}$$

$$F(x) = \sqrt{|x|} + 10 \text{ (model B);}$$

$$F(x) = |x|^{3/4} + 10 \text{ (model C).}$$

Two cases of error process  $(\sigma(X_n)\varepsilon_{n+1})_{n \geq 0}$  are considered for each model:  $\sigma(x) = 1$  and  $\sigma(x) = \exp(-|x|)$ ,  $(\varepsilon_{n+1})_{n \geq 0}$  being independently and identically distributed  $N(0, \rho)$  with  $\rho = 1$  in model A and  $\rho = 0.5$  in models B and C (hence  $\alpha_0 = 0$  in all cases).

One sample size  $n = 100$  is investigated and we predict the values of  $H = 5$  observations. For each combination of model and  $\sigma$ , 50 independent sets of data are generated.

Tables 2.1 - 2.4 give the means and the standard deviation of these 50 errors  $EK$  and  $EM$  obtained for each model.

$a$	Means and standard deviations ( $\times 100$ ) of	
	$EK$	$EM$
0.8	2.18 (1.01)	1.64 (0.5)
0.9	1.70 (1.09)	0.84 (0.29)
1	9.45 (0.09)	1.02 (0.04)
1.02	10.22 (0.035)	2.29 (0.011)

*Table 2.1. Model A with  $\sigma(x) = 1$*   
(The standard deviations are in parentheses)

$a$	Means and standard deviations ( $\times 100$ ) of	
	$EK$	$EM$
0.8	0.28 ( $6.68 \times 10^{-3}$ )	$1.32 \times 10^{-4}$ ( $2.68 \times 10^{-6}$ )
0.9	1.11 (0.01)	0.035 ( $3.41 \times 10^{-4}$ )
1	9.47 ( $9.52 \times 10^{-3}$ )	1.02 ( $1.03 \times 10^{-3}$ )
1.02	10.22 ( $3.51 \times 10^{-3}$ )	2.29 ( $7.87 \times 10^{-4}$ )

*Table 2.2. Model A with  $\sigma(x) = \exp(-|x|)$*

$\sigma$	Means and standard deviations ( $\times 100$ ) of	
	$EK$	$EM$
1	2.87 (0.92)	2.88 (0.89)
$\exp(- x )$	$8.33 \times 10^{-4}$ ( $1.18 \times 10^{-4}$ )	$3.22 \times 10^{-6}$ ( $1.05 \times 10^{-6}$ )

*Table 2.3. Model B.*

$\sigma$	Means and standard deviations ( $\times 100$ ) of	
	$EK$	$EM$
1	2.30 (0.66)	2.31 (0.66)
$\exp(- x )$	$9.84 \times 10^{-4}$ ( $5.40 \times 10^{-4}$ )	$3.07 \times 10^{-7}$ ( $1.54 \times 10^{-7}$ )

*Table 2.4. Model C.*

One remarkable aspect of the study is how our procedure performed when the linear model is asymptotically stationary and unstable (tables 2.1 and 2.2,  $a = 0.9, 1, 1.02$ ).

For models B and C (tables 2.3 and 2.4), the error is insensitive to the choice of the predictor when  $\sigma(x) = 1$  but the difference between the two errors  $EK$  and  $EM$  are significant when  $\sigma(x) = \exp(-|x|)$ . In all cases, it is preferable to choose the  $L_1$ -method that gives estimators of the errors with smaller standard deviations.



### 2.3 Data Analysis

To confirm the advantages of the  $L_1$ -method, we shall apply it to a time series consisting of the IBM common stock closing prices (see Figure 2.1, Box and Jenkins [4]). We suppose that the model is such that  $\alpha_0 = 0$ .

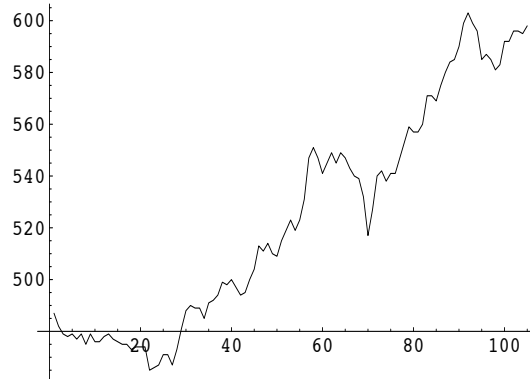


Figure 1: Datas from 6th June to 3rd September 1961

We choose two sample sizes :  $n = 90$  (datas from 6th June to 3rd September 1961) and  $n = 105$  (datas from 6th June to 18th September 1961).

For  $n = 90$ , the number of predictions  $H$  is equal to 5 and for  $n = 105$ ,  $H = 20$ .

	$EK (\times 100)$	$EM (\times 100)$
$n = 90$	2.76	0.79
$n = 105$	1.67	0.96

Table 2.5.

Table 2.5 shows that the error is smaller using  $L_1$ -method. However, the error obtained with the classical method of nonparametric regression is reduced choosing the sample size equal to 105. This phenomenon has already been noted in the model A above (with  $a = 1$  against  $a = 0.8$ ). Indeed, one remark that according to figure 1, the 80 first datas are observations of a process with unit root (the trend is an increasing straight line) and the 20 last are observations of a stationary process.

The three curves (Figure 2) represent the 20 last observations and their predictions according to the two methods.

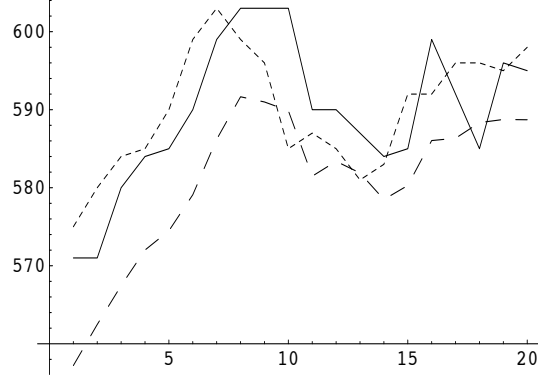


Figure 2: Observations and predictions

(The solid line represents the observations; the dotted line, the predictions by the  $L_1$ -method; the dashed line, the predictions by the classical method.)

### 3. Proofs

For simplicity of the proofs, we let  $S \subset \mathbb{R}^d$  be a compact set and we assume in this Section that  $\sup_S \|F\| \leq 1$ ,  $\sup_S \sigma \leq 1$  and  $\sigma$ ,  $F$  are Lipschitz functions (hence  $H_1 = H_2 = 1$ ) of order  $\leq 1$ , i.e:

$$\|F(x) - F(y)\| \leq \|x - y\|, |\sigma(x) - \sigma(y)| \leq \|x - y\|, \forall x, y \in \mathbb{R}^d.$$

We also assume that  $\sup_{n \geq 1} E[\|X_n\|] < \infty$ . According to Kronecker's Lemma, we then have w.p. 1:

$$\frac{1}{n(\log n)^2} \sum_{i=1}^n \|X_i\| \rightarrow 0, \text{ as } n \rightarrow \infty. \quad (3.0)$$

First of all, we state three lemmas, delaying their proofs to the end of this section.

**Lemma 3.1** *Let  $x \in \mathbb{R}^d$ . The probability measure  $\nu^x$  has a unique  $L_1$ -median which is  $F(x) + \sigma(x)\alpha_0$ .*

**Lemma 3.2** *If  $ad(4+d) < 1$  then w.p. 1:*

$$\sup_{x \in S} \sup_{n \geq 1} \|m_n(x)\| < \infty.$$

**Lemma 3.3** *If  $ad(4+d) < 1$  then w.p. 1:*

$$\sup_{x \in S} |\varphi(m_n(x)|x) - \varphi(F(x) + \sigma(x)\alpha_0|x)| \rightarrow 0, \text{ as } n \rightarrow \infty.$$

In the following, we denote by  $\psi$  the application defined for all  $\alpha \in \mathbb{R}^d$  by:

$$\psi(\alpha) = E[|\varepsilon - \alpha|].$$

Recall that we have:

$$\text{Argmin}_{\alpha \in \mathbb{R}^d} \psi(\alpha) = \{\alpha_0\}.$$

**Proof of Theorem 1.1** W.p. 1, for all  $\eta > 0$  and  $n \geq 1$ , there exists  $x_n \in S$  such that:

$$\sup_{x \in S} \|m_n(x) - (F(x) + \sigma(x)\alpha_0)\| \leq \|m_n(x_n) - (F(x_n) + \sigma(x_n)\alpha_0)\| + \eta. \quad (3.1)$$

Let us introduce the event:

$$\mathcal{B}_0 = [\sup_n \|m_n(x_n)\| < \infty, |\varphi(m_n(x_n)|x_n) - \varphi(F(x_n) + \sigma(x_n)\alpha_0|x_n)| \rightarrow 0].$$

By Lemmas 3.2 and 3.3, we have  $P(\mathcal{B}_0) = 1$ . Fix  $\omega \in \mathcal{B}_0$ . From any subsequence, one can extract a subsequence  $(n_k)_{k \geq 1}$  (depending on  $\omega$ ) such that:

$$x_{n_k} \rightarrow x_\infty \text{ and } m_{n_k}(x_{n_k})(\omega) \rightarrow \chi.$$

We first prove that  $\chi = F(x_\infty) + \sigma(x_\infty)\alpha_0$ . Recall that for all  $\alpha, x \in \mathbb{R}^d$ :

$$\varphi(\alpha|x) = E[|F(x) + \sigma(x)\varepsilon - \alpha|] = \sigma(x)\psi(\sigma(x)^{-1}(\alpha - F(x))). \quad (3.2)$$

Then, by continuity:

$$\varphi(\chi|x_\infty) = \lim_k \varphi(m_{n_k}(x_{n_k})(\omega)|x_{n_k}),$$

and since  $\omega \in \mathcal{B}_0$ , we have:

$$\varphi(\chi|x_\infty) = \lim_k \varphi(F(x_{n_k}) + \sigma(x_{n_k})\alpha_0|x_{n_k}).$$

Then, according to Lemma 3.1 and formula (3.2):

$$\begin{aligned}
\varphi(\chi|x_\infty) &= \lim_k \inf_{\alpha \in \mathbb{R}^d} \varphi(\alpha|x_{n_k}) \\
&= \lim_k \sigma(x_{n_k}) \inf_{\alpha \in \mathbb{R}^d} \psi(\sigma(x_{n_k})^{-1}(\alpha - F(x_{n_k}))) \\
&= \sigma(x_\infty) \inf_{\alpha \in \mathbb{R}^d} \psi(\alpha) = \sigma(x_\infty)\psi(\alpha_0) \\
&= \varphi(F(x_\infty) + \sigma(x_\infty)\alpha_0|x_\infty).
\end{aligned}$$

By Lemma 3.1,  $F(x_\infty) + \sigma(x_\infty)\alpha_0$  is the unique  $L_1$ -median of  $\nu^{x_\infty}$  hence  $\chi = F(x_\infty) + \sigma(x_\infty)\alpha_0$ . Moreover, by continuity of  $F$  and  $\sigma$ :

$$\|m_{n_k}(x_{n_k})(\omega) - (F(x_{n_k}) + \sigma(x_{n_k})\alpha_0)\| \rightarrow 0, \text{ as } k \rightarrow \infty.$$

As a conclusion, from any subsequence of the sequence:

$$(\|m_n(x_n)(\omega) - (F(x_n) + \sigma(x_n)\alpha_0)\|)_{n \geq 1},$$

one can extract a subsequence which converges to 0. Hence this sequence vanishes as  $n \rightarrow \infty$ . Since  $\omega \in \mathcal{B}_0$ , one deduce from (3.1) that w.p. 1:

$$\sup_{x \in S} \|m_n(x) - (F(x) + \sigma(x)\alpha_0)\| \rightarrow 0, \text{ as } n \rightarrow \infty,$$

hence the theorem  $\square$

We now prove Lemmas 3.1-3.3. The proof of Lemma 3.1 is easy.

**Proof of Lemma 3.1** By assumption on  $\varepsilon$ :

$$\text{Argmin}_{\alpha \in \mathbb{R}^d} \psi(\alpha) = \{\alpha_0\}.$$

By above and (3.2), we deduce that:

$$\text{Argmin}_{\alpha \in \mathbb{R}^d} \varphi(\alpha|x) = \{F(x) + \sigma(x)\alpha_0\},$$

i.e.  $\nu^x$  has a unique  $L_1$ -median  $F(x) + \sigma(x)\alpha_0$   $\square$

For notational simplicity, we shall write for all  $i \geq 1$  and  $u, x \in \mathbb{R}^d$ :

$$K_{i,x}(u) = i^{ad} K(i^a \|u - x\|).$$

**Lemma 3.4** *The stationary distribution  $\mu$  of the Markov chain  $(X_n)_{n \geq 0}$  is absolutely continuous with respect to the Lebesgue measure. If we denote by*

$k$  the density, we have  $I(S) = \inf_S k > 0$  and moreover, if  $2ad < 1$ , w.p. 1, for all  $n$  large enough and  $x \in S$ :

$$nI(S) \leq 2 \sum_{i=1}^n K_{i,x}(X_{i-1}).$$

**Proof** We deduce from an easy argument that the stationary distribution is absolutely continuous with respect to the Lebesgue measure (see Duflo [12], p. 187). Moreover, the transition probability of the Markov chain has a density  $p$  which is easily seen to be:

$$p(x, y) = \sigma(x)^{-d} g(\sigma(x)^{-1}(y - F(x))), \forall x, y \in \mathbb{R}^d.$$

Since  $\inf \sigma > 0$  and  $g$  is Lipschitz and bounded, we deduce from Proposition 7.1.8 by Duflo [12] that w.p. 1:

$$\sup_{x \in S} \left| \frac{1}{n} \sum_{i=1}^n K_{i,x}(X_{i-1}) - k(x) \right| \rightarrow 0, \text{ as } n \rightarrow \infty,$$

and hence one only needs now to prove that  $I(S) = \inf_S k > 0$ . By Duflo ([12], page 187), the density  $k$  satisfies:

$$k(y) = \int k(x) \sigma(x)^{-d} g(\sigma(x)^{-1}(y - F(x))) dx, \forall y \in \mathbb{R}^d. \quad (3.3)$$

Assume that there exists a compact set  $M \subset \mathbb{R}^d$  such that  $\inf_M k = 0$ . Then by (3.3), we have:

$$\begin{aligned} 0 = \inf_M k &\geq \int k(x) \sigma(x)^{-d} \inf_{y \in M} g(\sigma(x)^{-1}(y - F(x))) dx \\ &\geq \inf_{x, y \in M} \sigma(x)^{-d} g(\sigma(x)^{-1}(y - F(x))) \int_M k(x) dx. \end{aligned}$$

Since  $F$  and  $\sigma$  are continuous, we deduce from the assumptions on  $g$  that:

$$\int_M k(x) dx = 0,$$

and hence that  $k = 0$  almost everywhere on  $M$ . Fix  $y_0 \in M$  such that  $k(y_0) = 0$ . By (3.3), we then have:

$$\int k(x) \sigma(x)^{-d} g(\sigma(x)^{-1}(y_0 - F(x))) dx = 0,$$

so that almost everywhere (recall that  $\inf \sigma$  is positive),

$$k(x)g(\sigma(x)^{-1}(y_0 - F(x))) = 0.$$

By positivity of  $g$ , we deduce that  $k = 0$  almost everywhere. Since  $k$  is a density, we have a contradiction  $\square$

If  $n \geq 2$ , we denote by  $\gamma_n$  the real number:

$$\gamma_n = \frac{1}{n^{a(d+1)}(\log n)^2}.$$

We now define a covering of  $S$ , which will be used several times. One can find an increasing sequence of integers  $(l_n)_{n \geq 1}$  such that  $\sup_n l_n \gamma_n^d < \infty$  and for all  $n \geq 1$ , there exists  $x_1^n, \dots, x_{l_n}^n \in \mathbb{R}^d$  satisfying:

$$\sup_{n \geq 1} \sup_{k \leq l_n} \|x_k^n\| < \infty \text{ and } S \subset \bigcup_{k=1}^{l_n} B(x_k^n, \gamma_n).$$

Without loss of generality, we assume throughout that for all  $n \geq 1$  and  $k = 1, \dots, l_n$  :  $x_k^n \in S$ .

**Lemma 3.5** *For all  $i \geq 1$ , let  $\chi_i : \mathbb{R}^d \rightarrow \mathbb{R}_+$  such that for all  $x \in \mathbb{R}^d$ ,*

$$\chi_i(x) \leq 1 + \|x\|.$$

*If  $ad(4+d) < 1$ , we have w.p. 1, as  $n \rightarrow \infty$ :*

$$\max_{k \leq l_n} \frac{1}{\sum_{i=1}^n K_{i,x_k^n}(X_{i-1})} \left| \sum_{i=1}^n (\chi_i(\varepsilon_i) - E[\chi_i(\varepsilon)]) K_{i,x_k^n}(X_{i-1}) \right| \rightarrow 0.$$

*If moreover  $E[\chi_i(\varepsilon)] \rightarrow 0$  as  $i \rightarrow \infty$ , then w.p. 1:*

$$\max_{k \leq l_n} \frac{1}{\sum_{i=1}^n K_{i,x_k^n}(X_{i-1})} \left| \sum_{i=1}^n \chi_i(\varepsilon_i) K_{i,x_k^n}(X_{i-1}) \right| \rightarrow 0.$$

**Proof** Note that by Lemma 3.4, one only needs to prove that w.p. 1:

$$\frac{1}{n} \max_{k \leq l_n} \sum_{i=1}^n (\chi_i(\varepsilon_i) - E[\chi_i(\varepsilon)]) K_{i,x_k^n}(X_{i-1}) \rightarrow 0, \text{ as } n \rightarrow \infty.$$

Let  $\beta > 0$  and  $n \geq 1$ . Then,

$$\begin{aligned}
& P(|\max_{k \leq l_n} \sum_{i=1}^n (\chi_i(\varepsilon_i) - E[\chi_i(\varepsilon)]) K_{i,x_k^n}(X_{i-1})| \geq n\beta) \\
& \leq \sum_{k=1}^{l_n} P(|\sum_{i=1}^n (\chi_i(\varepsilon_i) - E[\chi_i(\varepsilon)]) K_{i,x_k^n}(X_{i-1})| \geq n\beta) \quad (3.4)
\end{aligned}$$

For all  $i \geq 1$ , let  $\mathcal{G}_i$  be the  $\sigma$ -field generated by  $X_0, \varepsilon_1, \dots, \varepsilon_i$  and  $\mathcal{G}_0$  be generated by  $X_0$ . By a martingale property, one may use the exponential inequality by Haeusler ([14], Lemma 1). If  $\|K\|_\infty$  denotes the supremum norm of  $K$ , we have for all  $n \geq 1$  and  $x \in S$ :

$$\begin{aligned}
& P(|\sum_{i=1}^n (\chi_i(\varepsilon_i) - E[\chi_i(\varepsilon)]) K_{i,x}(X_{i-1})| \geq n\beta) \\
& \leq n^{-2} + \sum_{i=1}^n P(|(\chi_i(\varepsilon_i) - E[\chi_i(\varepsilon)]) K_{i,x}(X_{i-1})| \geq \beta n (\log n)^{-1}) \\
& \quad + 2P(\sum_{i=1}^n E[(\chi_i(\varepsilon_i) - E[\chi_i(\varepsilon)])^2 K_{i,x}^2(X_{i-1}) | \mathcal{G}_{i-1}] \geq n^2 \beta^2 \exp(-3) (\log n)^{-1}) \\
& \leq n^{-2} + \sum_{i=1}^n P(|\chi_i(\varepsilon) - E[\chi_i(\varepsilon)]| n^{ad} \|K\|_\infty \geq \beta n (\log n)^{-1}) \\
& \quad + 2 \sum_{i=1}^n P(\text{var}(\chi_i(\varepsilon)) n^{2ad+1} \|K\|_\infty^2 \geq n^2 \beta^2 \exp(-3) (\log n)^{-1}) \\
& \leq n^{-2} + \sum_{i=1}^n P(|\chi_i(\varepsilon) - E[\chi_i(\varepsilon)]| \|K\|_\infty \geq \beta n^{1-ad} (\log n)^{-1}) \\
& \leq c (\log n)^3 n^{-2+3ad} + n^{-2},
\end{aligned}$$

where  $c > 0$  is a constant, since  $\varepsilon$  has a moment of order 3. Finally, Lemma 3.5 is then a straightforward consequence of (3.4) (recall that  $ad(4+d) < 1$  and  $\sup_n l_n \gamma_n^d < \infty$ )  $\square$

For all  $\lambda > 0$  and  $D \subset \mathbb{R}^d$ , let:

$$D^\lambda = \{y \in \mathbb{R}^d : \exists z \in D \text{ with } \|y - z\| \leq \lambda\}.$$

If moreover  $n \geq 1$ , we let:

$$\begin{aligned}
H(x, D) &= \int \|y\| I_D(y) \nu^x(dy); \\
H_n(x, D) &= \int \|y\| I_D(y) \nu_n^x(dy).
\end{aligned}$$

**Lemma 3.6** *Let  $\lambda > 0$  and assume that for some  $R \subset \mathbb{R}^d$ ,  $\nu^x(R^c) \leq \lambda$  and  $H(x, R^c) \leq \lambda$  for all  $x \in S$ . Then, if  $ad(4+d) < 1$ , we have w.p. 1:*

$$\begin{aligned} \limsup_n \sup_{x \in S} \nu_n^x((R^\lambda)^c) &\leq \lambda; \\ \limsup_n \sup_{x \in S} H_n(x, (R^\lambda)^c) &\leq \lambda. \end{aligned}$$

**Proof** We only prove the second inequality. Let  $n \geq 1$  and  $x \in B(x_k^n, \gamma_n)$ , for some  $k = 1, \dots, l_n$ . An easy calculation shows that w.p. 1:

$$\begin{aligned} &H_n(x, (R^\lambda)^c) \\ &\leq |H_n(x, (R^\lambda)^c) - H_n(x_k^n, (R^\lambda)^c)| + H_n(x_k^n, (R^\lambda)^c) \\ &\leq \frac{\|K\|_L \sum_{i=1}^n \|X_i\|}{(\log n)^2 \sum_{i=1}^n K_{i,x}(X_{i-1})} + \frac{n\|K\|_L}{(\log n)^2 \sum_{i=1}^n K_{i,x}(X_{i-1})} H_n(x_k^n, (R^\lambda)^c) \\ &\quad + H_n(x_k^n, (R^\lambda)^c). \end{aligned}$$

According to (3.0), the sequence of random variables

$$Z_n = \frac{1}{n(\log n)^2} \sum_{i=1}^n \|X_i\|$$

vanishes w.p. 1. By above and Lemma 3.4, we deduce that w.p. 1 and for all  $n$  large enough:

$$\begin{aligned} \sup_{x \in S} H_n(x, (R^\lambda)^c) &\leq \frac{2Z_n \|K\|_L}{I(S)} \\ &\quad + \left(1 + \frac{2\|K\|_L}{I(S)(\log n)^2}\right) \max_{k \leq l_n} H_n(x_k^n, (R^\lambda)^c). \end{aligned} \quad (3.5)$$

One only needs now to study the asymptotic behavior of the sequence  $(\max_{k \leq l_n} H_n(x_k^n, (R^\lambda)^c))_n$ . Since  $K$  has a compact support contained in  $[0, 1]$ , we have for all  $x \in S$  and  $n \geq 1$ :

$$\begin{aligned} &\sum_{i=1}^n \|X_i\| I_{\{X_i \in (R^\lambda)^c\}} K_{i,x}(X_{i-1}) \\ &= \sum_{i=1}^n \|X_i\| I_{\{F(X_{i-1}) + \sigma(X_{i-1})\varepsilon_i \in (R^\lambda)^c, \|X_{i-1} - x\| > i^{-a}\}} K_{i,x}(X_{i-1}) \\ &\quad + \sum_{i=1}^n \|X_i\| I_{\{F(X_{i-1}) + \sigma(X_{i-1})\varepsilon_i \in (R^\lambda)^c, \|X_{i-1} - x\| \leq i^{-a}\}} K_{i,x}(X_{i-1}) \end{aligned}$$



$$\begin{aligned}
&\leq \sum_{i=1}^n (1 + \|\varepsilon_i\|) i^{-a} K_{i,x}(X_{i-1}) \\
&\quad + \sum_{i=1}^n \|F(x) + \sigma(x)\varepsilon_i\| I_{\{F(x)+\sigma(x)\varepsilon_i \in R^c, \|\varepsilon_i\| \leq \lambda i^a - 1\}} K_{i,x}(X_{i-1}), \\
&\quad + \sum_{i=1}^n \|F(x) + \sigma(x)\varepsilon_i\| I_{\{\|\varepsilon_i\| > \lambda i^a - 1\}} K_{i,x}(X_{i-1}), \\
&\leq \sum_{i=1}^n (1 + \|\varepsilon_i\|) i^{-a} K_{i,x}(X_{i-1}) \\
&\quad + \sum_{i=1}^n \|F(x) + \sigma(x)\varepsilon_i\| I_{\{F(x)+\sigma(x)\varepsilon_i \in R^c\}} K_{i,x}(X_{i-1}), \\
&\quad + \sum_{i=1}^n (1 + \|\varepsilon_i\|) I_{\{\|\varepsilon_i\| > \lambda i^a - 1\}} K_{i,x}(X_{i-1}).
\end{aligned}$$

By Lemma 3.5, we have w.p. 1,

$$\lim_n \max_{k \leq l_n} H_n(x_k^n, (R^\lambda)^c) \leq \sup_{x \in S} E[\|F(x) + \sigma(x)\varepsilon\| I_{\{F(x)+\sigma(x)\varepsilon \in R^c\}}].$$

Finally, by (3.5), we deduce that w.p. 1:

$$\lim_n \sup_{x \in S} H_n(x, (R^\lambda)^c) \leq \sup_{x \in S} H(x, R^c),$$

hence the lemma  $\square$

**Lemma 3.7** *Let  $\alpha \in \mathbb{R}^d$ . Then, if  $ad(4+d) < 1$ , we have w.p. 1:*

$$\sup_{x \in S} |\varphi_n(\alpha|x) - \varphi(\alpha|x)| \rightarrow 0, \text{ as } n \rightarrow \infty.$$

**Proof** As in the beginning of the proof of Lemma 3.6, one deduce from an easy calculation and Lemma 3.4 that w.p. 1, for all  $n$  large enough:

$$\begin{aligned}
&\sup_{x \in S} |\varphi_n(\alpha|x) - \varphi(\alpha|x)| \\
&\leq \gamma_n + \frac{2\|K\|_L}{I(S)(\log n)^2} \left( \frac{1}{n} \sum_{i=1}^n \|X_i\| + \|\alpha\| \right) + \frac{2\|K\|_L}{I(S)(\log n)^2} \max_{k \leq l_n} \varphi_n(\alpha|x_k^n) \\
&\quad + \max_{k \leq l_n} |\varphi_n(\alpha|x_k^n) - \varphi(\alpha|x_k^n)|.
\end{aligned}$$

By (3.0) and since  $\varphi(\alpha|\cdot)$  is finite over  $S$ , one only needs to prove that w.p. 1:

$$\max_{k \leq l_n} |\varphi_n(\alpha|x_k^n) - \varphi(\alpha|x_k^n)| \rightarrow 0, \text{ as } n \rightarrow \infty.$$

For all  $n \geq 1$  and  $x \in S$ , since  $K$  has a compact support contained in  $[0, 1]$ , we have for all  $n$ :

$$\begin{aligned} & \sum_{i=1}^n K_{i,x}(X_{i-1}) |\varphi_n(\alpha|x) - \varphi(\alpha|x)| \\ & \leq \left| \sum_{i=1}^n (\|X_i - \alpha\| - \|F(x) + \sigma(x)\varepsilon_i - \alpha\|) K_{i,x}(X_{i-1}) I_{\{\|X_{i-1}-x\| \leq i^{-a}\}} \right| \\ & \quad + \left| \sum_{i=1}^n (\|X_i - \alpha\| - \|F(x) + \sigma(x)\varepsilon_i - \alpha\|) K_{i,x}(X_{i-1}) I_{\{\|X_{i-1}-x\| > i^{-a}\}} \right| \\ & \quad + \left| \sum_{i=1}^n (\|F(x) + \sigma(x)\varepsilon_i - \alpha\| - E[\|F(x) + \sigma(x)\varepsilon - \alpha\|]) K_{i,x}(X_{i-1}) \right| \\ & \leq \sum_{i=1}^n (1 + \|\varepsilon_i\|) i^{-a} K_{i,x}(X_{i-1}) \\ & \quad + \left| \sum_{i=1}^n (\|F(x) + \sigma(x)\varepsilon_i - \alpha\| - E[\|F(x) + \sigma(x)\varepsilon - \alpha\|]) K_{i,x}(X_{i-1}) \right|. \end{aligned}$$

Lemma 3.7 is then a straightforward consequence of Lemma 3.5  $\square$

**Lemma 3.8** *Assume that  $ad(4+d) < 1$ . Then, w.p. 1 and for all  $u > 0$ :*

$$\sup_{x \in S} \sup_{\|\alpha\| \leq u} |\varphi_n(\alpha|x) - \varphi(\alpha|x)| \rightarrow 0, \text{ as } n \rightarrow \infty.$$

**Proof** For all  $\|\alpha\| \leq u$ , we denote by  $f_\alpha$  the function:

$$f_\alpha(y) = \|y - \alpha\|, \quad y \in \mathbb{R}^d,$$

and by  $\mathcal{A}$  the set  $\mathcal{A} = \{f_\alpha, \|\alpha\| \leq u\}$ . Note that the family  $(\nu^x)_{x \in S}$  is tight and  $\varepsilon$  is integrable, so that for all  $\lambda > 0$ , there exists a bounded set  $R \subset \mathbb{R}^d$  such that  $\forall x \in S, \nu^x(R^c) \leq \lambda$  and  $H(x, R^c) \leq \lambda$ . By Ascoli, the set  $\mathcal{A}|_R$  (restriction of functions in  $\mathcal{A}$  defined on  $R$ ) endowed with the uniform topology is totally bounded. Consequently, one can find  $\{g_1, \dots, g_p\} \subset \mathcal{A}$  such that for all  $\|\alpha\| \leq u$ , there exists  $j = 1, \dots, p$  with:

$$\sup_{y \in R} |f_\alpha(y) - g_j(y)| \leq \lambda.$$

But  $g_j \in \mathcal{A}$ , hence:

$$\sup_{y \in R^\lambda} |f_\alpha(y) - g_j(y)| \leq \sup_{y \in R} |f_\alpha(y) - g_j(y)| + 2\lambda \leq 3\lambda.$$

Then w.p. 1 and for all  $x \in S$ ,  $n \geq 1$ :

$$\begin{aligned} & \sup_{\|\alpha\| \leq u} |\varphi_n(\alpha|x) - \varphi(\alpha|x)| \\ &= \sup_{\|\alpha\| \leq u} |\nu_n^x(f_\alpha) - \nu^x(f_\alpha)| \\ &\leq u\nu_n^x((R^\lambda)^c) + u\nu^x((R^\lambda)^c) + H_n(x, (R^\lambda)^c) + H(x, (R^\lambda)^c) \\ &\quad + \sup_{\|\alpha\| \leq u} |\nu_n^x(f_\alpha I_{R^\lambda}) - \nu^x(f_\alpha I_{R^\lambda})| \\ &\leq u\nu_n^x((R^\lambda)^c) + u\nu^x((R^\lambda)^c) + H_n(x, (R^\lambda)^c) + H(x, (R^\lambda)^c) \\ &\quad + 6\lambda + \max_{j \leq p} |\nu_n^x(g_j I_{R^\lambda}) - \nu^x(g_j I_{R^\lambda})| \\ &\leq 2u\nu_n^x((R^\lambda)^c) + 2u\nu^x((R^\lambda)^c) + 2H_n(x, (R^\lambda)^c) + 2H(x, (R^\lambda)^c) \\ &\quad + 6\lambda + \max_{j \leq p} |\nu_n^x(g_j) - \nu^x(g_j)|. \end{aligned}$$

Since  $R \subset R^\lambda$ , we deduce from Lemmas 3.6 and 3.7 that for all  $\lambda > 0$ , w.p. 1:

$$\limsup_n \sup_{x \in S} \sup_{\|\alpha\| \leq u} |\varphi_n(\alpha|x) - \varphi(\alpha|x)| \leq (10 + 4u)\lambda.$$

Letting  $\lambda \rightarrow 0$  through the rationals, one obtains that w.p. 1:

$$\sup_{x \in S} \sup_{\|\alpha\| \leq u} |\varphi_n(\alpha|x) - \varphi(\alpha|x)| \rightarrow 0, \text{ as } n \rightarrow \infty.$$

It is then a classical exercise to prove that the null set may be chosen so as to be independent of  $u > 0$   $\square$

**Proof of Lemma 3.2** By Lemma 3.8 and since  $F$  and  $\sigma$  are bounded by 1 over  $S$ , w.p. 1, there exists  $N \geq 1$  such that for all  $n \geq N$  and  $x \in S$ :

$$\varphi_n(0|x) \leq 1 + \varphi(0|x) \leq 2 + E[|\varepsilon|].$$

Consequently, w.p. 1:

$$\sup_{n \geq N} \sup_{x \in S} \varphi_n(0|x) < \infty.$$

Moreover, for all  $n = 1, \dots, N$  and  $x \in S$ :

$$\varphi_n(0|x) \leq \max_{i=1, \dots, N} \|X_i\|.$$

We deduce from the above inequalities that w.p. 1:

$$\sup_{n \geq 1} \sup_{x \in S} \varphi_n(0|x) < \infty. \quad (3.6)$$

Now, observe that if  $n \geq 1$  and  $x \in S$ :

$$\|m_n(x)\| \leq \varphi_n(m_n(x)|x) + \varphi_n(0|x),$$

and obviously:

$$\varphi_n(m_n(x)|x) \leq \varphi_n(0|x),$$

so that:

$$\sup_{n \geq 1} \sup_{x \in S} \|m_n(x)\| \leq 2 \sup_{n \geq 1} \sup_{x \in S} \varphi_n(0|x),$$

which is almost surely finite according to (3.6)  $\square$

Our last task is to prove Lemma 3.3.

**Proof of Lemma 3.3** According to Lemma 3.2, the random variable

$$u = \sup_{x \in S} \|F(x) + \sigma(x)\alpha_0\| + \sup_{x \in S} \sup_{n \geq 1} \|m_n(x)\|$$

is almost surely finite. By Lemma 3.1, we then have for all  $n \geq 1$ :

$$\begin{aligned} & \sup_{x \in S} |\varphi(m_n(x)|x) - \varphi(F(x) + \sigma(x)\alpha_0|x)| \\ & \leq \sup_{x \in S} |\varphi(m_n(x)|x) - \varphi_n(m_n(x)|x)| \\ & \quad + \sup_{x \in S} \left| \inf_{\|\alpha\| \leq u} \varphi_n(\alpha|x) - \inf_{\|\alpha\| \leq u} \varphi(\alpha|x) \right| \\ & \leq 2 \sup_{x \in S} \sup_{\|\alpha\| \leq u} |\varphi_n(\alpha|x) - \varphi(\alpha|x)|, \end{aligned}$$

and the rightmost term vanishes almost surely according to Lemma 3.8  $\square$

**Acknowledgements** The authors are grateful to Alain Berline for its valuable comments.

## References

[1] Berline, A., Cadre, B., Gannoun A.: On the conditional  $L_1$ -median and its estimation, to appear in *J. Nonparametr. Statist.* (2001).

- [2] Berlinet, A., Cadre, B., Gannoun A.: Estimation of the conditional  $L_1$ -median from dependent observations, to appear in *Statist. Probab. Lett.* (2001).
- [3] Bosq, D.: *Linear processes in function spaces: theory and applications*, Lectures Notes in Statistics **149**, Springer-Verlag, New-York, 2000.
- [4] Box G., Jenkins G.: *Time series analysis: forecasting and control (Revised edition)*, Holden-Day, 1976.
- [5] Brockwell P.J., Davis R.A.: *Time series: theory and methods (Second edition)*, Springer Series in Statistics, Springer-Verlag, New-York, 1996.
- [6] Cadre, B.: Convergent estimators for the  $L_1$ -median of a Banach valued random variable, to appear in *Statistics* (2001).
- [7] Cadre, B.: Hölder properties of the  $L_1$ -median. Rate of convergence in the nonparametric estimation of the conditional  $L_1$ -median (mixing case), submitted.
- [8] Diebolt, J.: Testing the functions defining a nonlinear autoregressive time series, *Stochastic Process. Appl.* **36** (1990), 85-106.
- [9] Diebolt, J., Ngatchou Wandji, J.: A nonparametric test for general first-order autoregressive model with a mixed term, *C.R.A.S.* **322** (1996), 577-582.
- [10] Doukhan, P.: *Mixing - Properties and examples*, Lectures Notes in Statistics **85**, Springer-Verlag, New-York, 1994.
- [11] Doukhan, P., Ghindès, M.: Estimations dans le processus  $X_{n+1} = f(X_n) + \varepsilon_{n+1}$ , *C.R.A.S.* **291** (1980), 61-64.
- [12] Duflo, M.: *Random iterative models*, Applications of Mathematics **37**, Springer-Verlag, New-York, 1997.
- [13] Gannoun, A., Cadre, B., Berlinet, A.: L'estimation non paramétrique de la médiane spatiale conditionnelle en vue de la prévision, *Actes du Colloque de la Société Mathématique Tunisienne*, Tabarka (1999), 46-55.
- [14] Haeusler, E.: An exact rate of convergence in the functional central limit theorem for special martingale differences arrays, *Z. W.* **65** (1984), 523-534.
- [15] Kemperman, J.H.D.: The median of finite measure of Banach space, in *Statistical data analysis based on the  $L_1$ -norm and related methods*, Y. Dodge, North-Holland, Amsterdam, 1987, 217-230.
- [16] Milasevic P., Ducharme G.R.: Uniqueness of the spatial median, *Ann. Statist.* **15** (1987), 1332-1333.
- [17] Robinson P.: Robust nonparametric regression, *In Robust and nonlinear time series analysis. Lecture Notes in Statist.* **26** (1984), 246-255.
- [18] Small, K.G.: A survey of multidimensional medians, *Internat. Statist. Rev.* **58** (1990), 263-277.