

## SEMIPARAMETRIC DECONVOLUTION WITH UNKNOWN NOISE VARIANCE

CATHERINE MATIAS<sup>1</sup>

**Abstract.** This paper deals with semiparametric convolution models, where the noise sequence has a Gaussian centered distribution, with unknown variance. Non-parametric convolution models are concerned with the case of an entirely known distribution for the noise sequence, and they have been widely studied in the past decade. The main property of those models is the following one: the more regular the distribution of the noise is, the worst the rate of convergence for the estimation of the signal's density  $g$  is [5]. Nevertheless, regularity assumptions on the signal density  $g$  improve those rates of convergence [15]. In this paper, we show that when the noise (assumed to be Gaussian centered) has a variance  $\sigma^2$  that is unknown (actually, it is always the case in practical applications), the rates of convergence for the estimation of  $g$  are seriously deteriorated, whatever its regularity is supposed to be. More precisely, the minimax risk for the pointwise estimation of  $g$  over a class of regular densities is lower bounded by a constant over  $\log n$ . We construct two estimators of  $\sigma^2$ , and more particularly, an estimator which is consistent as soon as the signal has a finite first order moment. We also mention as a consequence the deterioration of the rate of convergence in the estimation of the parameters in the nonlinear errors-in-variables model.

**Mathematics Subject Classification.** 62G05, 62G07, 62C20.

### 1. INTRODUCTION

Consider a convolution model  $Y = X + \varepsilon$ , where the signal  $X$  has an unknown density  $g$  with respect to the Lebesgue measure on  $\mathbb{R}$ , and the error measurement  $\varepsilon$  is supposed to be Gaussian, centered, with variance  $\sigma^2$ , and independent of  $X$ . Deconvolution density estimation has been studied in depth by several authors. Recent related work include [1–5, 11, 16, 17, 21]. When  $\sigma^2$  is known, Fan [5] proved that for all fixed point  $x_0$  in  $\mathbb{R}$ ,  $g(x_0)$  can be approximated at the optimal rate of convergence  $(\log n)^{-(m+\alpha)/2}$ , when  $g$  is supposed to belong to the set

$$\mathcal{C}_{m,\alpha,\beta} = \{g \in L^1 : g \geq 0; \text{ and } \forall x \in \mathbb{R}, \forall \delta > 0, |g^{(m)}(x) - g^{(m)}(x + \delta)| \leq \beta\delta^\alpha\} \quad (1)$$

where  $m$  in  $\mathbb{N}$ ,  $\beta > 0$  and  $0 < \alpha \leq 1$  are known constants. More precisely,

$$\liminf_{n \rightarrow \infty} \inf_{\hat{T}_n} \sup_{g \in \mathcal{C}_{m,\alpha,\beta}} (\log n)^{(m+\alpha)} \mathbb{E}[\hat{T}_n - g(x_0)]^2 > 0,$$

---

*Keywords and phrases:* Convolution, deconvolution, density estimation, mixing distribution, normal mean mixture model, semiparametric mixture model, noise, variance estimation, minimax risk.

<sup>1</sup> UMR C 8628 du CNRS, Équipe de Probabilités, Statistique et Modélisation, bâtiment 425, Université Paris-Sud, 91405 Orsay Cedex, France; e-mail: [Catherine.Matias@math.u-psud.fr](mailto:Catherine.Matias@math.u-psud.fr)

and this rate of convergence is attained. Results about convergence in  $L^p$ -norm were also obtained in [4] and [6], where it is proved:

$$\liminf_{n \rightarrow \infty} \inf_{\hat{T}_n} \sup_{g \in \mathcal{C}_{m,\alpha,\beta}} (\log n)^{(m+\alpha)/2} \mathbb{E} \|\hat{T}_n - g\|_p > 0,$$

and this rate is attained. This result is improved in [15], when the density  $g$  is super-smooth. Denote by  $g^*$  the Fourier transform of  $g$ , and assume that the density  $g$  belongs to the set

$$\mathcal{SS}_{\alpha,\nu,\rho}(A) = \left\{ g \in L^1 : \int |g^*(t)|^2 (t^2 + 1)^\alpha \exp(2\rho|t|^\nu) dt \leq A \right\}, \tag{2}$$

for some positive constants  $\alpha, \nu, \rho$  and  $A$ . Pensky and Vidakovic [15] constructed an estimator  $\hat{g}_n$  of  $g$ , whose mean integrated square error

$$\text{MISE}(\hat{g}_n) = \mathbb{E} \int (\hat{g}_n(x) - g(x))^2 dx \tag{3}$$

satisfies

$$\sup_{g \in \mathcal{SS}_{\alpha,\nu,\rho}(A)} \text{MISE}(\hat{g}_n) = \begin{cases} O(n^{-\eta}(\log n)^\xi), & \text{if } \nu < 2, \\ O((\log n)^{-\alpha} \exp(-\zeta(\log n)^{\nu/2})) & \text{if } \nu \geq 2, \end{cases}$$

where  $\eta, \xi$  and  $\zeta$  have explicit forms. So that assuming the density of the signal super-smooth ensures faster rates of convergence, in the case of an entirely known noise density.

The question that naturally arises is what happens when  $\sigma^2$  is unknown? This problem becomes a particular case of a semiparametric model [19], and more precisely, of mixture models [10], known as the normal mean mixture model. This problem of measurements being contaminated with errors is used in many different areas such as physics or biology (practical problems of deconvolution can be found in [13]).

Semiparametric mixture models are studied in [9]. The author shows a loss of information for the finite-dimensional parameter when the model is constrained to allow for only discrete mixtures. In the normal mean mixture model, allowing discrete mixtures to have limit points leads to a breakdown of the classical  $\sqrt{n}$  inference for the finite-dimensional parameter. Can we be more precise in quantifying this breakdown? The answer is yes, and we will see that for regular mixtures, the estimation of the finite-dimensional parameter  $\sigma$  happens at a slower rate than  $(\log n)^{-1}$ .

Assuming that the error density is perfectly known seems to be unrealistic in many practical applications. In [14] a scheme is given to estimate  $\sigma^2$  when observations of the noise sequence are available. Our estimator of  $\sigma^2$  is based only on the observations of the convolution model. More precisely, we assume that we observe:

$$Y_n = X_n + \varepsilon_n ; \quad n \in \mathbb{N},$$

where  $\{X_n\}_{n \geq 0}$  is a sequence of independent and identically distributed random variables on  $\mathbb{R}$ , and  $\{\varepsilon_n\}_{n \geq 0}$  is a centered Gaussian white noise, with variance  $\sigma^2$ . The sequences  $\{X_n\}_{n \geq 0}$  and  $\{\varepsilon_n\}_{n \geq 0}$  are supposed to be independent. We denote by  $g$  (resp.  $h$ ) the density of the distribution of  $X$  (resp.  $Y$ ) with respect to the Lebesgue measure on  $\mathbb{R}$ . In this paper,  $\Phi_\sigma$  denotes the density of a Gaussian centered random variable, with variance  $\sigma^2$ , and the notation  $g * \Phi_\sigma$  stands for the convolution product between  $g$  and  $\Phi_\sigma$ . A density function  $g$  (with respect to the Lebesgue measure on  $\mathbb{R}$ ) is said to have “no Gaussian component” if the equality  $g = g' * \Phi_\sigma$  where  $g'$  is a density function on  $\mathbb{R}$  implies  $\sigma = 0$  (and then  $g = g'$ ). We consider:

$$\mathcal{G} = \{g \in L^1 : g \text{ density without Gaussian component}\},$$

the set of densities for the signal, with respect to the Lebesgue measure on  $\mathbb{R}$ . Note that the restriction on  $g$  enables us to identify the parameter  $\sigma^2$ . Denote also

$$\mathcal{H} = \{g * \Phi_\sigma : g \in \mathcal{G} \text{ and } \sigma > 0\},$$

the set of densities for the observed sequence, with respect to the Lebesgue measure on  $\mathbb{R}$ . We denote by  $\mathbb{P}_{\sigma,g}$  the distribution of the observations  $\{Y_n\}_{n \geq 0}$  and  $\mathbb{E}_{\sigma,g}$  the corresponding expectation.

When  $\sigma^2$  is unknown, we prove that the pointwise estimation of  $g$  is deteriorated in many cases. In fact, the minimax quadratic risk over a set of regular densities, for the estimation of  $\sigma^2$  is lower bounded by a constant divided by  $\log n$ . Then, the estimation of  $g(x_0)$  (for some fixed point  $x_0$  in  $\mathbb{R}$ ) will not be possible, uniformly over a set of regular densities, at a faster rate than  $(\log n)^{-1}$ .

This paper is organized as follows. Section 2 gives lower bounds for the estimations of  $\sigma^2$  and  $g(x_0)$  in the convolution model with unknown variance for the noise sequence, using the van Trees inequality [8]. In Section 3, we construct an estimator of  $\sigma$  that is consistent, assuming nothing but a first order moment on  $g$ . This section also gives the rate of convergence of an estimator of  $\sigma^2$  when the Laplace transform of  $g$  has some decrease at infinity. Section 4 gives the main examples for which the lower bound results apply. It also gives a result about the degradation of the estimation of linear functionals coming from polynomial functions in the errors-in-variables model. Proofs are given in Section 5.

## 2. LOWER BOUNDS

We give lower bounds for the minimax quadratic risks for the estimation of  $\sigma$  and  $g(x_0)$ , when  $x_0$  is a fixed point in  $\mathbb{R}$ :

$$\inf_{\hat{T}_n} \sup_{\sigma \in \mathcal{V}(\sigma_0)} \sup_{g \in \mathcal{R}} \left[ \mathbb{E}_{\sigma,g}(\hat{T}_n - \sigma^2)^2 \right]^{1/2} \quad \text{and} \quad \inf_{\hat{T}_n} \sup_{\sigma \in \mathcal{V}(\sigma_0)} \sup_{g \in \mathcal{R}} \left[ \mathbb{E}_{\sigma,g}(\hat{T}_n - g(x_0))^2 \right]^{1/2},$$

where  $\hat{T}_n$  ranges over all estimators based on the observations  $Y_1, \dots, Y_n$ ,  $\mathcal{R}$  is a set of densities with properties to be precised, and  $\mathcal{V}(\sigma_0)$  is a neighborhood of a fixed point  $\sigma_0 > 0$ . We use the van Trees inequality [8] on a suited one-dimensional sub-model. In fact, the difficulty of the model is contained in a “worst-case” sub-model, that is to say a worst one-dimensional sub-model. As always in this kind of proof, the purpose is to exhibit densities  $h_t$  and  $h_0$  which are close to each other, in the sense that the Fisher information  $I(t)$  of the model is small for fixed  $t > 0$ , whereas the corresponding parameters  $\sigma_t^2$  and  $\sigma_0^2$  are well-separated.

For some density probability  $g_0$  and  $\tau > 0$ , define the set

$$\mathcal{H}(g_0, \tau) = \{g_0 * (\Phi_{\sqrt{t}} \mathbb{1}_{\tau| \cdot | \leq \sqrt{t}}) ; 0 \leq t \leq \tau^2\}, \tag{4}$$

with the convention that the centered Gaussian density with variance zero is the dirac mass at the point zero.

**Assumption 1.** *The set of densities  $\mathcal{H}(g_0, \tau)$  is included in  $\mathcal{G}$  for some positive  $\tau$ .*

**Assumption 2.** *The function  $\alpha_0(y) = \int_{-1}^1 g_0(y + u) du$  satisfies  $\int [\alpha_0(y)]^{-1} e^{-y^2/2} dy < \infty$ .*

**Assumption 3.** *The function  $g_0$  is three times continuously differentiable with  $\sup_{x \in \mathbb{R}^*} \left| \frac{g'_0(x)}{x} \right| < +\infty$ ,  $g''_0(0) \neq 0$  and  $\|g_0^{(3)}\|_\infty < +\infty$ .*

These assumptions are satisfied for example with  $g_0(x) = \pi^{-1}/(1 + x^2)$ .

Fix the points  $x_0$  in  $\mathbb{R}$  and  $\sigma_0$  in  $\mathbb{R}^+$ , a real positive parameter  $\tau$ , a bounded function  $g_0$  satisfying Assumptions 1 and 2. For all  $0 < t \leq \tau^2$ , we define

$$p_t(u) = \mathcal{C}_\tau \Phi_{\sqrt{t}\sigma_0}(u) \mathbb{1}_{\tau|u| \leq \sqrt{t}\sigma_0},$$

where the normalizing constant  $\mathcal{C}_\tau$  is equal to  $(\int \Phi_1(z) \mathbb{1}_{\tau|z| \leq 1} dz)^{-1}$ . Now, let us construct a path in  $\mathcal{G}$ :

$$\forall 0 < t \leq \tau^2, \quad h_t = g_0(\cdot - x_0) * p_t * \Phi_{\sqrt{1-t}\sigma_0} \quad \text{and} \quad h_0 = g_0(\cdot - x_0) * \Phi_{\sigma_0}. \tag{5}$$

The function  $p_t$  is the truncated density of a centered Gaussian random variable, with variance  $t\sigma_0^2$ . Since the convolution of  $\Phi_{\sqrt{t}\sigma_0}$  with  $\Phi_{\sqrt{1-t}\sigma_0}$  is equal to  $\Phi_{\sigma_0}$ , we can expect that the densities  $h_t$  (for  $t > 0$ ) and  $h_0$  are close to each other (in a sense to be precised).

**Theorem 2.1.** *Assume that  $g_0$  is a bounded function satisfying Assumptions 1 and 2 and fix  $\tau > 0$ . Then for all  $\sigma_0 > 0$  and every neighborhood  $\mathcal{V}(\sigma_0)$  of  $\sigma_0$ , we have*

$$\liminf_{n \rightarrow \infty} \inf_{\hat{T}_n} \sup_{\sigma \in \mathcal{V}(\sigma_0)} \sup_{g \in \mathcal{H}(g_0, \tau)} (\log n) \left[ \mathbb{E}_{\sigma, g}(\hat{T}_n - \sigma^2)^2 \right]^{1/2} > 0, \tag{6}$$

where the infimum is taken over all estimators  $\hat{T}_n$  based on the observations  $Y_1, \dots, Y_n$ .

A slight adaptation of this path gives the corresponding result on the pointwise estimation of  $g$  at  $x_0$ .

**Theorem 2.2.** *Assume that  $g_0$  is a bounded function satisfying Assumptions 1, 2 and 3 and fix  $\tau > 0$ . Then for all real number  $x_0$ , for all  $\sigma_0 > 0$ , and every neighborhood  $\mathcal{V}(\sigma_0)$  of  $\sigma_0$ , we have*

$$\liminf_{n \rightarrow \infty} \inf_{\hat{T}_n} \sup_{\sigma \in \mathcal{V}(\sigma_0)} \sup_{g \in \mathcal{H}(g_0, \tau)} (\log n) \left[ \mathbb{E}_{\sigma, g}(\hat{T}_n - g(x_0))^2 \right]^{1/2} > 0,$$

where the infimum is taken over all estimators  $\hat{T}_n$  based on the observations  $Y_1, \dots, Y_n$ .

A consequence of these theorems is that the minimax risk for the estimation of  $\sigma^2$  or  $g(x_0)$ , over a class of densities containing the set  $\mathcal{H}(g_0, \tau)$  for some well-chosen density  $g_0$  and some  $\tau > 0$  small enough, is lower bounded by a constant over  $\log n$ . That means that the estimation of  $\sigma^2$  or  $g(x_0)$  happens at a slower rate than  $(\log n)^{-1}$  over these sets of densities. In particular, if  $\mathcal{R}$  is some ‘‘regular’’ set of densities without Gaussian component and  $g_0$  is chosen in  $\mathcal{R}$ , then  $\mathcal{R}$  contains the path  $\mathcal{H}(g_0, \tau)$  for some small enough  $\tau > 0$ . Section 4 gives the main examples of sets  $\mathcal{R}$  for which these theorems apply.

### 3. ESTIMATION OF THE NOISE VARIANCE

In this part, we propose two different procedures for the estimation of the noise variance  $\sigma^2$ . The first one is the most powerful as it gives an estimator consistent as soon as we assume a first order moment on the signal. The second procedure applies when the Laplace transform of the density of the signal  $g$  has some decrease at infinity, and is interesting as its rate of convergence is explicit. In particular, it includes the case of densities  $g$  with a fixed support  $[-M; M]$ .

We denote by  $g^*$  (resp.  $h^*$ ) the Fourier transform of  $g$  (resp.  $h$ ) and fix the noise variance  $\sigma^2 > 0$ . For all  $\zeta$  in  $\mathbb{R}$  and all  $\tau > 0$ , define

$$\alpha(\zeta; \tau) = g^*(\zeta) e^{-\zeta^2(\sigma^2 - \tau^2)/2} = h^*(\zeta) e^{\zeta^2 \tau^2/2}. \tag{7}$$

The function  $\alpha$  is the product of the Fourier transform of  $h$  (the density of the distribution of the observations) and the function  $(\Phi_\tau^*)^{-1}$ . When  $\tau$  equals the standard deviation  $\sigma$ , the function  $\alpha(\cdot; \sigma)$  equals  $g^*$ , the characteristic function of the signal. Note also that since  $g$  has no Gaussian component, the function  $\zeta \mapsto g^*(\zeta) e^{\zeta^2 v^2/2}$ , for a fixed  $v > 0$ , is not the Fourier transform of a positive measure. We observe the following properties of  $\alpha$ :

- when  $\tau \leq \sigma$ , the function  $\alpha(\cdot; \tau)$  is the Fourier transform of the positive measure  $g * \Phi_{\sqrt{\sigma^2 - \tau^2}}$ ;

- when  $\tau > \sigma$ , the function  $\alpha(\cdot; \tau)$  is no longer a Fourier transform of a probability measure on the real line (since  $g$  is supposed to belong to the set  $\mathcal{G}$  and then has no Gaussian component). By Bochner's theorem (see for example [7]), the function  $\alpha(\cdot; \tau)$  is then not positive definite.

So that we conclude:

$$\sigma = \inf\{\tau ; \alpha(\cdot, \tau) \text{ is not positive definite}\}.$$

In the rest of the section,  $u$  is used to denote a point in  $\mathbb{C}^p$ , where  $p$  may change along the lines, and  $\|u\|$  denotes the norm  $(\sum_{j=1}^p |u_j|^2)^{1/2}$ . The idea is that the real parameter  $\sigma$  is the first value of  $\tau$  for which the function  $\alpha(\cdot; \tau)$  is not positive definite. That means that  $\sigma$  is the smallest value of  $\tau$  satisfying that there exists an integer  $n$  and a  $n$ -tuple of real numbers  $\{t_k; 1 \leq k \leq n\}$ , such that the smallest eigenvalue of the matrix  $(\alpha(t_k - t_l; \tau))_{1 \leq k, l \leq n}$  is negative.

$$\sigma = \inf \left\{ \tau ; \exists n \in \mathbb{N}, \exists (t_k)_{1 \leq k \leq n} \in \mathbb{R} : \inf_{u; \|u\|=1} \sum_{1 \leq k, l \leq n} u_k \alpha(t_k - t_l; \tau) \bar{u}_l \leq 0 \right\}.$$

We approximate the function  $\alpha$  by its empirical estimator  $\hat{\alpha}_n$ :

$$\hat{\alpha}_n(\zeta; \tau) = \left( \frac{1}{n} \sum_{p=1}^n e^{i\zeta Y_p} \right) e^{\zeta^2 \tau^2 / 2}. \tag{8}$$

Let us construct a dense family  $\{t_{k,n}\}_{n \geq 0}$  of points in  $\mathbb{R}$ . Consider  $(k_n)_{n \geq 0}$  and  $(\ell_n)_{n \geq 0}$  two sequences of numbers increasing to infinity, in such a way that  $\ell_n/k_n$  also increases to infinity. The points  $t_{k,n} = k/k_n$  form a partition of the interval  $[-\ell_n/k_n; \ell_n/k_n]$  when the integer  $k$  ranges from  $-\ell_n$  to  $\ell_n$ . We define

$$\hat{\sigma}_n = \inf \left\{ \tau ; \inf_{u; \|u\|=1} \sum_{-\ell_n \leq k, l \leq \ell_n} u_k \hat{\alpha}_n(t_{k,n} - t_{l,n}; \tau) \bar{u}_l < -\epsilon_n \right\} \tag{9}$$

where  $(\epsilon_n)_{n \geq 0}$  is a sequence of positive numbers decreasing to zero. This estimator is computed considering the matrices  $T_n(\tau) = \{\hat{\alpha}_n(t_{k,n} - t_{l,n}; \tau)\}_{-\ell_n \leq k, l \leq \ell_n}$ . The graph of the function  $\tau \mapsto \lambda_{\min}(T_n(\tau))$ , where  $\lambda_{\min}(T)$  denotes the smallest eigenvalue of the matrix  $T$ , gives the value of  $\hat{\sigma}_n$  by considering the first value of  $\tau$  such that  $\lambda_{\min}(T_n(\tau)) < -\epsilon_n$ .

**Assumption 4.** Fix  $\Sigma > 0$  and choose the parameters  $\ell_n; k_n$  and  $\epsilon_n$  in the following way:

- $\frac{\ell_n}{k_n} = \frac{1}{\Sigma} \sqrt{\frac{a \log n}{2}}$  for some  $0 < a < 1/2$ ;
- $k_n = n^{1/2-a-b}$  for some  $b > 0$  and  $2a + b < 1/2$ ;
- $\epsilon_n = \frac{\sqrt{2a}}{\Sigma} \frac{\sqrt{\log n}}{n^b} v_n$  where  $v_n$  is an increasing sequence of numbers converging to infinity in such a way that  $\epsilon_n$  converges to zero, and  $v_n^{-1} = o((\log \log n)^{-1/2})$ .

**Theorem 3.1.** For all  $\Sigma > 0$ , under Assumption 4, and for all  $\sigma \in ]0; \Sigma]$ , all  $g$  in  $\mathcal{G}$  such that  $\int |x|g(x)dx < \infty$ , we have:

$$\forall \epsilon > 0, \lim_{n \rightarrow \infty} \mathbb{P}_{\sigma, g}(|\hat{\sigma}_n - \sigma| \geq \epsilon) = 0.$$

We only specify the asymptotic behaviour of the parameters  $\ell_n, k_n, \epsilon_n$  and  $v_n$ . Their values should be justified by empirical applications, but this is beyond the scope of this paper.

Now, we propose another estimator of  $\sigma$ , assuming that the distribution of the signal possesses a Laplace transform not rapidly increasing at infinity. This framework contains for example the case of  $g$  having a support included in some fixed compact set. The construction of this estimator of  $\sigma$  is based on the behaviour of the

Laplace transform of  $g$  at infinity: as the distribution of the signal has no Gaussian component, its Laplace transform should increase lower than  $e^{Ct^2}$  at infinity. Assume that the density of the signal  $g$  belongs to the following set of functions:

$$\mathcal{L}_v = \left\{ g \in L^1 / g \geq 0, \int g(x)dx = 1, \left| \log \left( \int e^{tx} g(x) dx \right) \right| \leq t^2 v(t) \right\}, \tag{10}$$

for some positive function  $v$  vanishing at infinity. Note for example that when  $v(t) = M/|t|$ , the set of functions  $\mathcal{L}_v$  contains the set of functions with support included in  $[-M; M]$ . In the rest of the paper, we will use the abbreviated notation:

$$\mathcal{L}(M) = \left\{ g \in L^1 / g \geq 0, \int g(x)dx = 1, \left| \log \left( \int e^{tx} g(x) dx \right) \right| \leq M|t| \right\}.$$

When the density  $g$  of the signal belongs to  $\mathcal{L}_{v,M}$  for some function  $v$  vanishing at infinity, the variance of the noise becomes identifiable, as can be seen by computing the Laplace transform of the observations. In fact, we have the equality

$$\forall g \in \mathcal{L}_{v,M}, \quad \sigma^2 = \lim_{t \rightarrow \infty} \frac{2}{t^2} \log \mathbb{E}_{\sigma,g} (e^{tY_1}).$$

It is then natural to define the following empirical estimator of  $\sigma$ :

$$\hat{\sigma}_{L,n}^2(t_n) = \frac{2}{t_n^2} \log \left( \frac{1}{n} \sum_{j=1}^n e^{t_n Y_j} \right),$$

where  $(t_n)_{n \geq 0}$  is some sequence of positive numbers increasing to infinity (the subscript  $L$  stands for Laplace).

**Proposition 3.2.** *For any  $\Sigma > 0$  and some  $0 < \alpha < 1$ , define  $t_n = \sqrt{\alpha \log n} / (2\Sigma)$ . Then for all positive function  $v$  vanishing at infinity:*

$$\sup_{\sigma \in [0; \Sigma]} \sup_{g \in \mathcal{L}_v} [\mathbb{E}_{\sigma,g} (\hat{\sigma}_{L,n}^2(t_n) - \sigma^2)]^{1/2} \leq \frac{8\Sigma^2}{\alpha(\log n)n^{(1-\alpha)/2}} + 2v(t_n).$$

Now, we consider the special case  $v(t) = M/|t|$ , and obtain as a straightforward application:

**Corollary 3.3.** *For all  $M > 0$ , all  $\Sigma > 0$  and some  $0 < \alpha < 1$ , define  $t_n = \sqrt{\alpha \log n} / (2\Sigma)$ . Then*

$$\sup_{\sigma \in [0; \Sigma]} \sup_{g \in \mathcal{L}(M)} [\mathbb{E}_{\sigma,g} (\hat{\sigma}_{L,n}^2(t_n) - \sigma^2)]^{1/2} \leq \frac{4M\Sigma}{\sqrt{\alpha \log n}}.$$

Note that Theorem 2.1 does not apply for the set  $\mathcal{L}(M)$  since any function  $g_0$  with compact support will not satisfy Assumption 2.

#### 4. EXAMPLES

We now give some examples of sets  $\mathcal{R}$  containing a bounded function  $g_0$  satisfying Assumptions 1, 2, 3, and such that  $\mathcal{R}$  contains the set  $\mathcal{H}(g_0, \tau)$  for some small enough  $\tau$ , and we compare our results with existing ones.

**Example 4.1.** Consider the set of functions  $\mathcal{C}_{m,\alpha,\beta}$  defined by (1). The regularity assumption on the function  $g$  has been used in [5] when the distribution of the noise sequence is known. Fan proved that the minimax rate of convergence when the density of the noise sequence is known and super smooth, is  $(\log n)^{-(m+\alpha)}$ . Adding

the condition on non-Gaussian components, in order to have an identifiable model, we prove that when  $\sigma$  is unknown, the rate of convergence in the pointwise estimation of  $g$  is seriously deteriorated, as it becomes slower than  $(\log n)^{-1}$ .

**Example 4.2.** Consider now the set of functions  $\mathcal{SS}_{\alpha,\nu,\rho}(A)$  defined by (2). Pensky and Vidakovic [15] gave the following improvement on the results of Fan: when the density  $g$  of the distribution of the signal is known to be super-smooth, they constructed an estimator of  $g$ , whose MISE error (3) is upper bounded by a constant divided by some power of  $n$ , times some power of  $\log n$  (see [15] for more details). But when the variance of the noise sequence is unknown, the fact that  $g$  is super-smooth does not improve the rate of convergence of its pointwise estimation, which remains slower than  $(\log n)^{-1}$ . It seems to be also the case for the MISE error in the estimation of  $g$ .

A consequence of these results concerns the non-linear errors-in-variables model. Consider the model where the observations  $\{(Y_n; Z_n)\}_{n \geq 0}$  satisfy the following relations:

$$\begin{cases} Z_n &= f_\beta(X_n) + \eta_n \\ Y_n &= X_n + \varepsilon_n \end{cases}, \quad \forall n \geq 0, \tag{11}$$

where the function  $f_\beta$  is known up to the finite dimensional parameter  $\beta$ , the errors  $\{(\eta_n; \varepsilon_n)\}_{n \geq 0}$  are independent, identically distributed and centered with respective variances  $\sigma_\eta^2$  and  $\sigma_\varepsilon^2 = 1$ , the variables  $\varepsilon_n$  being Gaussian and the sequence  $\{X_n\}_{n \geq 0}$  is not observed and is a sequence of independent and identically distributed random variables with distribution admitting a density  $g$  with respect to the Lebesgue measure on  $\mathbb{R}$ . The purpose is to estimate the parameter  $\beta$  in this model where  $g$  is considered as a nuisance. Taupin [18] constructed an estimator of  $\beta$  based on the estimation of the conditional expectation  $\mathbb{E}(f_\beta(X_n)|Y_n)$ . The fact is that this conditional expectation writes in the following form:

$$\mathbb{E}(f_\beta(X_n)|Y_n = y) = \frac{\int f_\beta(x)g(x)\Phi_1(x - y)dx}{\int g(x)\Phi_1(x - y)dx}.$$

Taupin [18] constructed an estimator based on the observations  $\{Y_n\}_{n \geq 0}$ , of the linear functional  $\Gamma_f$  of the density  $g$ , defined by the formula:

$$\Gamma_f(y) = \int f(x)g(x)\Phi_1(x - y)dx; \quad \forall y \in \mathbb{R}.$$

When  $f$  is identically equal to one, the functional  $\Gamma_f$  equals to the density  $h$  of the observations  $Y_n$ . Rates of convergence for this estimator of the functionals when  $f$  is either a polynomial function or a trigonometric function of the form  $x \mapsto \sum_{j=0}^\ell \beta_j \cos(jx)$  or of the form  $x \mapsto \sum_{j=0}^\ell \beta_j \sin(jx)$  for some integer  $\ell$  and real fixed parameters  $(\beta_j)_{0 \leq j \leq \ell}$  are given in [18] and are shown to be minimax in [12]. Typically, the minimax rate of convergence in  $L^p$ -norm or in pointwise quadratic risk for a functional  $\Gamma_f$  where  $f$  is a polynomial function of degree less or equal to  $\ell$  is equal to  $(\log n)^{(2\ell+1)/4}/\sqrt{n}$  [18]. In the case of the estimation of  $h$ , those rates of convergence are not deteriorated when  $\sigma^2$  becomes unknown. Now consider the case where  $f$  is a polynomial function with degree  $\ell$  greater or equal to one. We prove that the estimation of  $\Gamma_f$  is seriously deteriorated in this case when  $\sigma^2$  is unknown as the minimax quadratic risk becomes lower bounded by a constant divided by  $\log n$ .

Let  $P$  be a polynomial function of degree  $\ell$  greater or equal to one, and  $g_0$  a density probability, and define the set

$$\mathcal{G}_P(g_0, \tau) = \left\{ \Gamma = \int P(x)\Phi_\sigma(\cdot - x)g(x)dx ; g = g_0 * (\Phi_{\sqrt{t}}\mathbb{1}_{\tau|\cdot| \leq \sqrt{t}}), 0 \leq t \leq \tau^2, \sigma > 0 \right\},$$

of those functionals constructed with a density  $g$  of the form  $g_0 * (\Phi_{\sqrt{t}} \mathbb{1}_{\tau|\cdot| \leq \sqrt{t}})$  for some  $0 \leq t \leq \tau^2$ . We have the following theorem concerning the estimation of a functional  $\Gamma$  in  $\mathcal{G}_P(g_0, \tau)$ .

**Theorem 4.3.** *Assume that  $g_0$  is a bounded function that is supposed to be  $\ell$  times continuously differentiable, satisfying Assumptions 1, 2 and 3 with  $\|g_0^{(\ell)}\|_\infty < \infty$ , and fix  $\tau > 0$ . Then, for all fixed real number  $y_0$  and every polynomial function  $P$  of degree  $\ell$  greater or equal to one, we have:*

$$\liminf_{n \rightarrow \infty} \inf_{\widehat{\Gamma}_n} \sup_{\Gamma \in \mathcal{G}_P(g_0, \tau)} (\log n) \left[ \mathbb{E}(\widehat{\Gamma}_n - \Gamma(y_0))^2 \right]^{1/2} > 0,$$

where the infimum is taken over all the estimators  $\widehat{\Gamma}_n$  based on the observations  $Y_1, \dots, Y_n$ .

### 5. PROOFS

*Proof of Proposition 2.1.* Without loss of generality, we assume that  $x_0 = 0$  and  $\sigma_0 = 1$ . The first thing to check is that the parametric path belongs to the model. But  $g_0$  is chosen so that for all  $0 \leq t \leq \tau^2$ , the density  $g_t = g_0 * p_t$  has no Gaussian component. By applying definition (5) of the density  $h_t$ , we can write, for all  $0 < t \leq \tau^2$ :

$$\begin{aligned} h_t(y) &= \mathcal{C}_\tau h_0(y) - (\mathcal{C}_\tau h_0 - h_t)(y) \\ &= \mathcal{C}_\tau h_0(y) - g_0 * (\mathcal{C}_\tau \Phi_1 - p_t * \Phi_{\sqrt{1-t}})(y). \end{aligned} \tag{12}$$

Note that we have the following identity:  $p_t = \mathcal{C}_\tau \Phi_{\sqrt{t}}(1 - \mathbb{1}_{\tau|\cdot| > \sqrt{t}})$ , so that

$$(\mathcal{C}_\tau \Phi_1 - p_t * \Phi_{\sqrt{1-t}})(y) = \mathcal{C}_\tau \left[ \Phi_1 - \left( \Phi_{\sqrt{t}} - \Phi_{\sqrt{t}} \mathbb{1}_{\{\tau|\cdot| > \sqrt{t}\}} \right) * \Phi_{\sqrt{1-t}} \right](y).$$

Since the convolution between the normal densities  $\Phi_{\sqrt{t}}$  and  $\Phi_{\sqrt{1-t}}$  is equal to  $\Phi_1$ , we have

$$(\mathcal{C}_\tau \Phi_1 - p_t * \Phi_{\sqrt{1-t}})(y) = \mathcal{C}_\tau \left( \Phi_{\sqrt{t}} \mathbb{1}_{\{\tau|\cdot| > \sqrt{t}\}} \right) * \Phi_{\sqrt{1-t}}(y).$$

Combining with (12), we get

$$\begin{aligned} h_t(y) &= \mathcal{C}_\tau h_0(y) - \mathcal{C}_\tau g_0 * \left( \Phi_{\sqrt{t}} \mathbb{1}_{\{\tau|\cdot| > \sqrt{t}\}} \right) * \Phi_{\sqrt{1-t}}(y) \\ &= \mathcal{C}_\tau h_0(y) - \mathcal{C}_\tau \iint \Phi_{\sqrt{1-t}}(y-u) \Phi_{\sqrt{t}}(u-v) \mathbb{1}_{\tau|u-v| > \sqrt{t}} g_0(v) dv du. \end{aligned} \tag{13}$$

We have:

$$\begin{aligned} \Phi_{\sqrt{t}}(u-v) \Phi_{\sqrt{1-t}}(y-u) &= \frac{1}{2\pi\sqrt{t(1-t)}} \exp \left[ -\frac{(u-(1-t)v-ty)^2}{2t(1-t)} - \frac{v^2}{2t} - \frac{y^2}{2(1-t)} + \frac{[(1-t)v+ty]^2}{2t(1-t)} \right] \\ &= \Phi_{\sqrt{t(1-t)}}(u-(1-t)v-ty) \Phi_1(v-y), \end{aligned}$$



and then returning to (13),

$$\begin{aligned}
 h_t(y) &= \mathcal{C}_\tau h_0(y) - \frac{\mathcal{C}_\tau}{\sqrt{2\pi}} \int \left[ \int \frac{1}{\sqrt{t(1-t)}} \Phi_1 \left( \frac{u - (1-t)v - ty}{\sqrt{t(1-t)}} \right) \mathbb{1}_{\tau|u-v| > \sqrt{t}} du \right] e^{-\frac{(v-y)^2}{2}} g_0(v) dv \\
 &= \mathcal{C}_\tau h_0(y) - \frac{\mathcal{C}_\tau}{\sqrt{2\pi}} \int \left[ \int_{\frac{(1/\tau) + \sqrt{t}(v-y)}{\sqrt{1-t}}}^{+\infty} \Phi_1(z) dz + \int_{-\infty}^{\frac{-(1/\tau) + \sqrt{t}(v-y)}{\sqrt{1-t}}} \Phi_1(z) dz \right] e^{-\frac{(v-y)^2}{2}} g_0(v) dv. \tag{14}
 \end{aligned}$$

This expression of  $h_t$  is useful in the computation of the corresponding Fisher information. In the rest of this section, the notation  $\mathbb{E}_t$  stands as an abbreviation for  $\mathbb{E}_{\sigma_t, g_t}$ . The Fisher information associated to our path is defined by:

$$I(t) = \mathbb{E}_t \left( \frac{\partial \log h_t}{\partial t}(Y) \right)^2 = \int \left( \frac{\partial h_t}{\partial t} \right)^2 (y) h_t^{-1}(y) dy, \quad \text{for all } 0 \leq t \leq \tau^2. \tag{15}$$

**Lemma 5.1.** *The Fisher information satisfies:*

$$\forall 0 \leq t \leq \tau^2, I(t) = I(0)(1 + o(1)) = \frac{\mathcal{C}_\tau^2}{4\pi^2} \tau^{-2} e^{-1/\tau^2} \left( \int f_0(y)^2 h_0(y)^{-1} dy \right) (1 + o(1)),$$

where  $f_0$  is defined by

$$f_0(y) = \int [1 - (v - y)^2] e^{-\frac{(v-y)^2}{2}} g_0(v) dv.$$

The proof of this lemma stands after this one.

Now, let us consider  $\lambda_0(t)dt$  a probability measure on  $[0; 1]$  satisfying the following conditions

- $\lambda_0(0) = \lambda_0(1) = 0$ ;
- $t \mapsto \lambda_0(t)$  is continuously differentiable in  $]0; 1[$ ;
- $\lambda_0(t)dt$  has finite Fisher information:

$$J_0 = \int_0^1 \frac{\lambda_0'(t)^2}{\lambda_0(t)} dt,$$

where the prime stands for derivation with respect to the parameter  $t$ .

Rescaling this measure on the interval  $[0; \tau^2]$  we define:

$$\lambda(t)dt = \frac{1}{\tau^2} \lambda_0 \left( \frac{t}{\tau^2} \right) dt$$

which has the Fisher information  $J_0/\tau^4$ . Denote by  $\mathbb{E}_\lambda(I)$  the quantity  $\int I(t)\lambda(t)dt$  ( $t$  is seen as a random variable with values in  $[0; \tau^2]$ , distributed according to the probability measure  $\lambda(t)dt$ ). The van Trees inequality [8] for the estimation of the variance of the noise in the convolution model gives us for small enough  $\tau$ :

$$\begin{aligned}
 \inf_{\hat{T}_n} \sup_{\sigma \in \mathcal{V}(\sigma_0)} \sup_{g \in \mathcal{H}(g_0, \tau)} \mathbb{E}_{\sigma, g} (\hat{T}_n - \sigma^2)^2 &\geq \inf_{\hat{T}_n} \int_0^{\tau^2} \mathbb{E}_t (\hat{T}_n - \sigma_t^2)^2 \lambda(t) dt \\
 &\geq \left( \int \frac{\partial \sigma_t^2}{\partial t}(t) \lambda(t) dt \right)^2 (n \mathbb{E}_\lambda(I) + J_0/\tau^4)^{-1}.
 \end{aligned}$$

And then

$$\inf_{\hat{T}_n} \sup_{\sigma \in \mathcal{V}(\sigma_0)} \sup_{g \in \mathcal{H}(g_0, \tau)} \mathbb{E}_{\sigma, g} (\hat{T}_n - \sigma^2)^2 \geq (n \mathbb{E}_\lambda(I) + J_0/\tau^4)^{-1}.$$

Using Lemma 5.1,

$$\mathbb{E}_\lambda(I) = I(0)(1 + o(1)) = \frac{\mathcal{C}_\tau^2}{4\pi^2} \tau^{-2} e^{-1/\tau^2} \left( \int f_0(y)^2 h_0(y)^{-1} dy \right) (1 + o(1)),$$

and then, returning to the van Trees inequality

$$\inf_{\hat{T}_n} \sup_{\sigma \in \mathcal{V}(\sigma_0)} \sup_{g \in \mathcal{H}(g_0, \tau)} \mathbb{E}_{\sigma, g} (\hat{T}_n - \sigma^2)^2 \geq \left[ n \frac{\mathcal{C}_\tau^2}{4\pi^2} \left( \int f_0(y)^2 h_0(y)^{-1} dy \right) \tau^{-2} e^{-1/\tau^2} (1 + o(1)) + J_0/\tau^4 \right]^{-1}.$$

Choosing  $\tau^{-1} = \sqrt{\log n}$ , we get:

$$\liminf_{n \rightarrow \infty} \inf_{\hat{T}_n} \sup_{\sigma \in \mathcal{V}(\sigma_0)} \sup_{g \in \mathcal{H}(g_0, \tau)} (\log n)^2 \mathbb{E}_{\sigma, g} (\hat{T}_n - \sigma^2)^2 > 0,$$

which achieves the proof of Proposition 2.1.  $\square$

*Proof of Lemma 5.1.* We calculate the derivative of  $h_t$  with respect to the parameter  $t$ , using equation (14):

$$\begin{aligned} \frac{\partial h_t}{\partial t}(y) &= -\frac{\mathcal{C}_\tau}{\sqrt{2\pi}} \int \left[ -\Phi_1 \left( \frac{(1/\tau) + \sqrt{t}(v-y)}{\sqrt{1-t}} \right) \left( \frac{(v-y)}{2\sqrt{t(1-t)}} + \frac{(1/\tau) + \sqrt{t}(v-y)}{2(1-t)^{3/2}} \right) \right. \\ &\quad \left. + \Phi_1 \left( \frac{-(1/\tau) + \sqrt{t}(v-y)}{\sqrt{1-t}} \right) \left( \frac{(v-y)}{2\sqrt{t(1-t)}} + \frac{-(1/\tau) + \sqrt{t}(v-y)}{2(1-t)^{3/2}} \right) \right] e^{-\frac{(y-v)^2}{2}} g_0(v) dv. \end{aligned}$$

But the equalities

$$\begin{aligned} \Phi_1 \left( \frac{(1/\tau) + \sqrt{t}(v-y)}{\sqrt{1-t}} \right) &= \frac{1}{\sqrt{2\pi}} \exp \left( -\frac{\tau^{-2} + t(v-y)^2}{2(1-t)} \right) \exp \left( -\frac{\sqrt{t}(v-y)}{\tau(1-t)} \right) \\ \Phi_1 \left( \frac{-(1/\tau) + \sqrt{t}(v-y)}{\sqrt{1-t}} \right) &= \frac{1}{\sqrt{2\pi}} \exp \left( -\frac{\tau^{-2} + t(v-y)^2}{2(1-t)} \right) \exp \left( +\frac{\sqrt{t}(v-y)}{\tau(1-t)} \right) \end{aligned}$$

lead to:

$$\begin{aligned} \frac{\partial h_t}{\partial t}(y) &= \frac{-\mathcal{C}_\tau}{2\pi} \int e^{-\frac{\tau^{-2} + t(v-y)^2}{2(1-t)}} \left[ \left( \frac{(v-y)}{2\sqrt{t(1-t)}} + \frac{-(1/\tau) + \sqrt{t}(v-y)}{2(1-t)^{3/2}} \right) e^{+\frac{\sqrt{t}(v-y)}{\tau(1-t)}} \right. \\ &\quad \left. - \left( \frac{(v-y)}{2\sqrt{t(1-t)}} + \frac{(1/\tau) + \sqrt{t}(v-y)}{2(1-t)^{3/2}} \right) e^{-\frac{\sqrt{t}(v-y)}{\tau(1-t)}} \right] e^{-\frac{(y-v)^2}{2}} g_0(v) dv. \end{aligned}$$

Simple calculations will give:

$$\begin{aligned} \frac{\partial h_t}{\partial t}(y) &= \frac{-\mathcal{C}_\tau}{2\pi(1-t)^{3/2}} \int (v-y) e^{-\frac{\tau^{-2} + t(v-y)^2}{2(1-t)}} \frac{1}{\sqrt{t}} \operatorname{sh} \left( \frac{\sqrt{t}(v-y)}{\tau\sqrt{1-t}} \right) e^{-\frac{(y-v)^2}{2}} g_0(v) dv \\ &\quad + \frac{\mathcal{C}_\tau}{2\pi(1-t)^{3/2}\tau} \int e^{-\frac{\tau^{-2} + t(v-y)^2}{2(1-t)}} \operatorname{ch} \left( \frac{\sqrt{t}(v-y)}{\tau\sqrt{1-t}} \right) e^{-\frac{(y-v)^2}{2}} g_0(v) dv. \end{aligned} \tag{16}$$

The second term in the right hand side of the last equality is well-defined when  $t$  is equal to zero. The first one is continuous at  $t = 0$ , observing that:

- $\frac{1}{\sqrt{t}} \operatorname{sh} \left( \frac{\sqrt{t}(v-y)}{\tau\sqrt{1-t}} \right) \xrightarrow{t \rightarrow 0} \frac{v-y}{\tau}$ ;
- using a Taylor formula:

$$\frac{1}{\sqrt{t}} \operatorname{sh} \left( \frac{\sqrt{t}(v-y)}{\tau\sqrt{1-t}} \right) = \frac{(v-y)}{2\sqrt{1-t}} \int_{-1/\tau}^{1/\tau} e^{\frac{x\sqrt{t}(v-y)}{\sqrt{1-t}}} dx,$$

we get an upper-bound, valid for all  $0 < t \leq \min(1/2; \tau^2)$ ,

$$\left| \frac{1}{\sqrt{t}} \operatorname{sh} \left( \frac{\sqrt{t}(v-y)}{\tau\sqrt{1-t}} \right) \right| \leq \frac{\sqrt{2}}{\tau} |v-y| e^{|v-y|}. \quad (17)$$

This leads to the domination, valid for all  $0 < t \leq \min(1/2; \tau^2)$ ,

$$\left| (v-y) e^{-\frac{\tau^{-2}+t(v-y)^2}{2(1-t)}} \frac{1}{\sqrt{t}} \operatorname{sh} \left( \frac{\sqrt{t}(v-y)}{\tau\sqrt{1-t}} \right) e^{-\frac{(y-v)^2}{2}} g_0(v) \right| \leq \frac{\sqrt{2}}{\tau} (v-y)^2 e^{-\tau^{-2}/2} e^{|v-y|} e^{-\frac{(y-v)^2}{2}} g_0(v)$$

and the dominating function, as a function of the variable  $v$ , belongs to  $L^1$ .

Then, dominated convergence combined with the expression (16) gives the continuity of  $[(\partial h_t)/(\partial t)](y)$  at  $t = 0$ :

$$\begin{aligned} \left( \frac{\partial h_t}{\partial t} \right)_{|t=0} (y) &= \frac{\mathcal{C}_\tau e^{-1/2\tau^2}}{2\pi \tau} \int [1 - (v-y)^2] e^{-\frac{(y-v)^2}{2}} g_0(v) dv = \frac{\mathcal{C}_\tau e^{-1/2\tau^2}}{2\pi \tau} f_0(y) \\ \text{where } f_0(y) &= \int [1 - (v-y)^2] e^{-\frac{(y-v)^2}{2}} g_0(v) dv \end{aligned}$$

belongs to  $L^1$ . Then, we have by definition

$$I(t) = \mathbb{E}_t \left( \frac{\partial \log h_t}{\partial t} (Y) \right)^2 = \int \left( \frac{\partial h_t}{\partial t} \right)^2 (y) h_t^{-1}(y) dy.$$

And now, we claim the continuity of  $t \mapsto I(t)$  at  $t = 0$ , using that:

- $t \mapsto \left( \frac{\partial h_t}{\partial t} \right)^2 (y) h_t^{-1}(y)$  is continuous at  $t = 0$  for all  $y \in \mathbb{R}$ ;
- consider the expression (16), and use the inequality  $|a+b|^2 \leq 2(a^2+b^2)$ , combined with the Cauchy-Schwarz inequality (recall that  $g_0$  is a probability density). We get the following domination:

$$\begin{aligned} \left| \frac{\partial h_t}{\partial t} (y) \right|^2 &\leq \frac{\mathcal{C}_\tau^2}{2\pi^2(1-t)^3} \int |v-y|^2 e^{-\frac{\tau^{-2}+t(v-y)^2}{(1-t)}} \frac{1}{t} \left| \operatorname{sh} \left( \frac{\sqrt{t}(v-y)}{\tau\sqrt{1-t}} \right) \right|^2 e^{-(v-y)^2} g_0(v) dv \\ &\quad + \frac{\mathcal{C}_\tau^2 \tau^{-2}}{2\pi^2(1-t)^3} \int e^{-\frac{\tau^{-2}+t(v-y)^2}{(1-t)}} \left| \operatorname{ch} \left( \frac{\sqrt{t}(v-y)}{\tau\sqrt{1-t}} \right) \right|^2 e^{-(v-y)^2} g_0(v) dv. \end{aligned}$$

Now, assume that  $0 < t \leq \min(1/2; \tau^2)$ , use the upper-bound (17) and the inequality  $|\operatorname{ch}(x)| \leq e^{|x|}$  to obtain:

$$\begin{aligned} \left| \frac{\partial h_t}{\partial t} (y) \right|^2 &\leq \frac{8\mathcal{C}_\tau^2 e^{-1/\tau^2}}{\pi^2 \tau^2} \int |v-y|^4 e^{2|v-y|} e^{-(v-y)^2} g_0(v) dv + \frac{4\mathcal{C}_\tau^2 e^{-1/\tau^2}}{\pi^2 \tau^2} \int e^{2|v-y|} e^{-(v-y)^2} g_0(v) dv \\ &\leq \frac{8\mathcal{C}_\tau^2 e^{-1/\tau^2}}{\pi^2 \tau^2} \int (1 + |v-y|^4) e^{2|v-y|} e^{-(v-y)^2} g_0(v) dv; \end{aligned} \quad (18)$$

- remember equality (14)

$$\begin{aligned}
 h_t(y) &= \mathcal{C}_\tau h_0(y) - \frac{\mathcal{C}_\tau}{\sqrt{2\pi}} \int \left[ \int_{\frac{\tau^{-1} + \sqrt{t}(v-y)}{\sqrt{1-t}}}^{+\infty} \Phi_1(z) dz + \int_{-\infty}^{\frac{-\tau^{-1} + \sqrt{t}(v-y)}{\sqrt{1-t}}} \Phi_1(z) dz \right] e^{-(v-y)^2/2} g_0(v) dv \\
 &= \frac{\mathcal{C}_\tau}{\sqrt{2\pi}} \int \left[ \int_{\frac{-\tau^{-1} + \sqrt{t}(v-y)}{\sqrt{1-t}}}^{\frac{\tau^{-1} + \sqrt{t}(v-y)}{\sqrt{1-t}}} \Phi_1(z) dz \right] e^{-(v-y)^2/2} g_0(v) dv.
 \end{aligned}$$

We lower bound this integral by its restriction to the set  $\{|v - y| \leq 1\}$  so that for small enough  $\tau$ , it can also be lower bounded by

$$h_t(y) \geq \frac{1}{2\sqrt{2\pi}} \int \mathbb{1}_{|v-y| \leq 1} e^{-(v-y)^2/2} g_0(v) dv,$$

and we obtain

$$h_t(y) \geq \frac{e^{-1/2}}{2\sqrt{2\pi}} \int_{-1}^1 g_0(y + u) du.$$

Now, by assumption, the function  $\alpha_0(y) = \int_{-1}^1 g_0(y + u) du$  satisfies

$$\int [\alpha_0(y)]^{-1} e^{-y^2/2} dy < \infty.$$

Combining with (18), we obtain a domination on the quantity

$$\left| \frac{\partial h_t}{\partial t}(y) \right|^2 h_t(y)^{-1}$$

by an integrable function of the variable  $y$ .

This achieves the proof of the continuity of the function  $t \rightarrow I(t)$  at the point  $t = 0$ . In conclusion, we have

$$I(t) = I(0)(1 + o(1)) = \frac{\mathcal{C}_\tau^2}{4\pi^2} \tau^{-2} e^{-1/\tau^2} \left( \int f_0(y)^2 h_0(y)^{-1} dy \right) (1 + o(1)). \quad \square$$

*Proof of Theorem 2.2.* Without loss of generality, we will assume  $x_0 = 0$  and  $\sigma_0 = 1$ . We use the same path as in the proof of Proposition 2.1, assuming moreover that  $g_0$  satisfies Assumption 3. The only thing to care about is the functional to be estimated. Here we have:

$$\begin{aligned}
 g_t(0) &= \int g_0(u) p_t(u) du = \mathcal{C}_\tau \int_{|v| \leq (1/\tau)} g_0(\sqrt{t}v) e^{-v^2/2} \frac{dv}{\sqrt{2\pi}} \\
 \frac{\partial g_t(0)}{\partial t} &= \frac{\mathcal{C}_\tau}{2\sqrt{t}} \int_{|v| \leq (1/\tau)} v g'_0(\sqrt{t}v) \Phi_1(v) dv.
 \end{aligned}$$

Expanding  $g'_0$  in the neighborhood of 0, there exists a point  $\bar{v}$  between 0 and  $\sqrt{t}v$  such that

$$g'_0(\sqrt{t}v) = g'_0(0) + \sqrt{t}v g''_0(0) + tv^2 g_0^{(3)}(\bar{v}).$$

Then, using the same notations as in the proof of Proposition 2.1,

$$\begin{aligned} \frac{\partial g_t(0)}{\partial t} &= g_0''(0) \frac{C_\tau}{2} \left( \int_{|v| \leq (1/\tau)} v^2 \Phi_1(v) dv \right) (1 + O(\sqrt{t})) \\ \left( \int \frac{\partial g_t(0)}{\partial t} (t) \lambda(t) dt \right)^2 &= [g_0''(0)]^2 \frac{C_\tau^2}{4} \left( \int_{|v| \leq (1/\tau)} v^2 \Phi_1(v) dv \right)^2 (1 + O(\tau)). \end{aligned}$$

Now, applying the van Trees inequality, we obtain that for small enough  $\tau$ :

$$\begin{aligned} \inf_{\hat{T}_n} \sup_{\sigma \in \mathcal{V}(\sigma_0)} \sup_{g \in \mathcal{H}(g_0, \tau)} \mathbb{E}_{\sigma, g} (\hat{T}_n - g(0))^2 &\geq \inf_{\hat{T}_n} \int_0^\tau \mathbb{E}_t (\hat{T}_n - g_t(0))^2 \lambda(t) dt \\ &\geq \left( \int \frac{\partial g_t(0)}{\partial t} (t) \lambda(t) dt \right)^2 (n \mathbb{E}_\lambda(I) + J_0/\tau^4)^{-1} \\ &\geq [g_0''(0)]^2 \frac{C_\tau^2}{4} \left( \int_{|v| \leq (1/\tau)} v^2 \Phi_1(v) dv \right)^2 (1 + o(1)) (n \mathbb{E}_\lambda(I) + J_0/\tau^4)^{-1}. \end{aligned}$$

Using Lemma 5.1, we have the equality

$$\mathbb{E}_\lambda(I) = I(0)(1 + o(1)) = \frac{C_\tau^2}{4\pi^2} \frac{e^{-1/\tau^2}}{\tau^2} \left( \int f_0(y)^2 h_0(y)^{-1} dy \right) (1 + o(1)).$$

And then, returning to the van Trees inequality

$$\inf_{\hat{T}_n} \sup_{\sigma \in \mathcal{V}(\sigma_0)} \sup_{g \in \mathcal{H}(g_0, \tau)} \mathbb{E}_{\sigma, g} (\hat{T}_n - g(0))^2 \geq \frac{[g_0''(0)]^2 \frac{C_\tau^2}{4} \left( \int_{|v| \leq (1/\tau)} v^2 \Phi_1(v) dv \right)^2 (1 + o(1))}{n \frac{C_\tau^2}{4\pi^2} \left( \int f_0(y)^2 h_0(y)^{-1} dy \right) \tau^{-2} e^{-1/\tau^2} (1 + o(1)) + J_0/\tau^4}.$$

Choosing  $\tau^{-1} = \sqrt{\log n}$ , we get

$$\liminf_{n \rightarrow \infty} \inf_{\hat{T}_n} \sup_{\sigma \in \mathcal{V}(\sigma_0)} \sup_{g \in \mathcal{H}(g_0, \tau)} (\log n)^2 \mathbb{E}_{\sigma, g} (\hat{T}_n - g(0))^2 > 0. \quad \square$$

*Proof of Theorem 3.1.* We have to choose the parameters to ensure that  $\hat{\sigma}_n$  will converge to the true value of the parameter  $\sigma$ . We will first study the quantity

$$\hat{\alpha}_n(t; \tau) - \alpha(t; \tau)$$

that represents the difference between the value of  $\alpha$  and its estimator. We have:

$$\hat{\alpha}_n(t; \tau) - \alpha(t; \tau) = e^{t^2 \tau^2 / 2} \frac{1}{\sqrt{n}} \mathbb{G}_n(f_t)$$

where  $\mathbb{G}_n$  is the empirical process associated to the observations (that is to say  $\mathbb{G}_n = \sqrt{n}(\mathbb{P}_n - \mathbb{P}_{\sigma, g})$  and  $\mathbb{P}_n$  is the empirical probability measure for the observations:  $\mathbb{P}_n(f) = (1/n) \sum_{i=1}^n f(Y_i)$ ), and  $f_t(u) = e^{itu}$ . This leads to the upper-bound:

$$\sup_{u: \|u\|=1} \sup_{\tau \in [0; \Sigma]} \left| \sum_{-\ell_n \leq k, l \leq \ell_n} u_k [\hat{\alpha}_n(t_{k,n} - t_{l,n}; \tau) - \alpha(t_{k,n} - t_{l,n}; \tau)] \bar{u}_l \right| \leq 2\ell_n \left( \sup_{|t| \leq \frac{\ell_n}{k_n}} |\mathbb{G}_n f_t| \right) \frac{e^{2\ell_n^2 \Sigma^2 / k_n^2}}{\sqrt{n}}.$$

We impose the following condition on the parameters

$$\frac{\ell_n}{k_n} = \frac{1}{\Sigma} \sqrt{\frac{a \log n}{2}} \quad \text{where } 0 < a < 1/2, \tag{19}$$

so that we get:  $e^{2\ell_n^2 \Sigma^2 / k_n^2} = n^a$ . It leads to

$$\sup_{u, \tau} \left| \sum_{-\ell_n \leq k, l \leq \ell_n} u_k [\hat{\alpha}_n(t_{k,n} - t_{l,n}; \tau) - \alpha(t_{k,n} - t_{l,n}; \tau)] \bar{u}_l \right| \leq \frac{\sqrt{2a} \sqrt{\log n k_n}}{\Sigma n^{1/2-a}} \left( \sup_{|t| \leq \frac{\ell_n}{k_n}} |\mathbb{G}_n f t| \right), \tag{20}$$

and we impose

$$k_n \leq n^{1/2-a-b}, \quad \text{where } b > 0,$$

so that the quantity

$$\frac{\sqrt{2a} \sqrt{\log n k_n}}{\Sigma n^{1/2-a}}$$

converges to zero. Recall that the parameter  $\epsilon_n$  is chosen as

$$\epsilon_n = \frac{\sqrt{2a} \sqrt{\log n k_n}}{\Sigma n^{1/2-a}} \times v_n,$$

where  $v_n$  is an increasing sequence of numbers converging to infinity such that  $\epsilon_n$  tends to zero:

$$v_n \xrightarrow{n \rightarrow \infty} \infty, \quad \epsilon_n \xrightarrow{n \rightarrow \infty} 0.$$

In fact, we choose the sequence  $(v_n)_{n \geq 0}$  increasing to infinity, under the constraint:

$$v_n = o\left(\frac{n^{1/2-a}}{\sqrt{\log n k_n}}\right).$$

This choice of the parameters is made in order to ensure the convergence to zero of the term:

$$\Delta_n = \mathbb{P}_{\sigma, g} \left( \sup_{u: \|u\|=1} \sup_{\tau \in [0; \Sigma]} \left| \sum_{-\ell_n \leq k, l \leq \ell_n} u_k [\alpha(t_{k,n} - t_{l,n}; \tau) - \hat{\alpha}(t_{k,n} - t_{l,n}; \tau)] \bar{u}_l \right| \geq \epsilon_n \right). \tag{21}$$

Let us prove this convergence. Using the domination (20), we have:

$$\Delta_n \leq \mathbb{P}_{\sigma, g} \left( \frac{\sqrt{2a} \sqrt{\log n k_n}}{\Sigma n^{1/2-a}} \left( \sup_{|t| \leq \frac{\ell_n}{k_n}} |\mathbb{G}_n f t| \right) \geq \epsilon_n \right).$$

So that,

$$\Delta_n \leq \mathbb{P}_{\sigma, g} \left( \sup_{|t| \leq \frac{\ell_n}{k_n}} |\mathbb{G}_n f t| \geq v_n \right) \leq \frac{1}{v_n} \mathbb{E}_{\sigma, g} \left( \sup_{|t| \leq \frac{\ell_n}{k_n}} |\mathbb{G}_n f t| \right). \tag{22}$$

Now we use a maximal inequality to control the mean of the empirical process. The following notations can be found in more details in [20]. We consider the class of functions  $\mathcal{F}_n$  defined by  $\{f_t; |t| \leq \ell_n/k_n\}$  (note that this class has an envelope function equal to one). The complexity of this family stands in its entropy defined through the bracketing numbers for this class. Theorem 2.7.11 in [20] applies in our context: denote by  $F(x) = 2|x|$  the function such that for all  $s, t$  in  $T_n = [-\ell_n/k_n; \ell_n/k_n]$

$$|f_t(x) - f_s(x)| = |e^{itx} - e^{isx}| \leq |s - t|F(x).$$

This theorem asserts that the bracketing numbers for the class  $\mathcal{F}_n$  (that means the minimal number of brackets of size  $\epsilon$  needed to cover  $\mathcal{F}_n$ ) are controlled by the covering numbers of  $T_n$  (*i.e.* the minimal number of balls of radius  $\epsilon$  needed to cover  $T_n$ ):

$$N_{[\cdot]}(\epsilon; \mathcal{F}_n; \|\cdot\|_{\mathbb{P}_{\sigma,g,2}}) \leq N\left(\frac{\epsilon}{2\|F\|_{\mathbb{P}_{\sigma,g,2}}}; T_n; |\cdot|\right)$$

(here  $\|\cdot\|_{\mathbb{P}_{\sigma,g,2}}$  denotes the  $L^2$ -norm under the measure  $\mathbb{P}_{\sigma,g}$ ). But it is easy to bound the covering numbers for  $T_n$

$$N\left(\frac{\epsilon}{2\|F\|_{\mathbb{P}_{\sigma,g,2}}}; T_n; |\cdot|\right) \leq \frac{4\ell_n\|F\|_{\mathbb{P}_{\sigma,g,2}}}{k_n\epsilon}.$$

So that we obtain the following control on the bracketing numbers for the class  $\mathcal{F}_n$

$$N_{[\cdot]}(\epsilon; \mathcal{F}_n; \|\cdot\|_{\mathbb{P}_{\sigma,g,2}}) \leq \frac{4\ell_n\|F\|_{\mathbb{P}_{\sigma,g,2}}}{k_n\epsilon}. \tag{23}$$

Let us define the entropy of this class by the formula:

$$J_{[\cdot]}(1; \mathcal{F}_n; \|\cdot\|_{\mathbb{P}_{\sigma,g,2}}) = \int_0^1 \sqrt{1 + \log N_{[\cdot]}(\epsilon; \mathcal{F}_n; \|\cdot\|_{\mathbb{P}_{\sigma,g,2}})} d\epsilon. \tag{24}$$

Now we apply Theorem 2.14.2 in [20]: there exist a universal constant  $C$  such that

$$\mathbb{E}_{\sigma,g} \left( \sup_{|t| \leq \frac{\ell_n}{k_n}} |\mathbb{G}_n f_t| \right) \leq C J_{[\cdot]}(1; \mathcal{F}_n; \|\cdot\|_{\mathbb{P}_{\sigma,g,2}}).$$

Combining with the definition of the entropy (24), with inequality (23) and using (19), we obtain that there exist some constant  $\kappa$  such that

$$\mathbb{E}_{\sigma,g} \left( \sup_{|t| \leq \frac{\ell_n}{k_n}} |\mathbb{G}_n f_t| \right) \leq \kappa (\log \log n)^{1/2}.$$

Now return to the quantity  $\Delta_n$  defined in (21) and to its upper-bound (22):

$$\Delta_n \leq \frac{1}{v_n} \mathbb{E}_{\sigma,g} \left( \sup_{|t| \leq \frac{\ell_n}{k_n}} |\mathbb{G}_n f_t| \right) \leq \frac{\kappa}{v_n} (\log \log n)^{1/2}. \tag{25}$$

Recall that the parameter  $v_n$  satisfies the constraint

$$v_n^{-1} = o\left((\log \log n)^{-1/2}\right),$$

in order to get the convergence of  $\Delta_n$  to zero:

$$\Delta_n \xrightarrow[n \rightarrow \infty]{} 0. \tag{26}$$

We will now prove the consistency of the estimator  $\hat{\sigma}_n$ . We start by studying what happens if  $\hat{\sigma}_n$  is greater than or equal to  $\sigma + \epsilon$  for some arbitrary  $\epsilon > 0$ . Using the definition of  $\hat{\sigma}_n$  (see (9))

$$\mathbb{P}_{\sigma,g}(\hat{\sigma}_n \geq \sigma + \epsilon) \leq \mathbb{P}_{\sigma,g} \left( \inf_{u: \|u\|=1} \sum_{-\ell_n \leq k, l \leq \ell_n} u_k \hat{\alpha}_n(t_{k,n} - t_{l,n}; \sigma + \epsilon) \bar{u}_l \geq -\epsilon_n \right).$$

But this quantity involves the estimator  $\hat{\alpha}$  whereas the real parameter  $\alpha$  is more tractable. Using the definition of  $\Delta_n$  (see(21)) we write:

$$\mathbb{P}_{\sigma,g}(\hat{\sigma}_n \geq \sigma + \epsilon) \leq \mathbb{P}_{\sigma,g} \left( \inf_{u: \|u\|=1} \sum_{-\ell_n \leq k, l \leq \ell_n} u_k \alpha(t_{k,n} - t_{l,n}; \sigma + \epsilon) \bar{u}_l \geq -2\epsilon_n \right) + \Delta_n. \tag{27}$$

But  $\Delta_n$  tends to zero as  $n$  tends to infinity. We will show that the first term in the right hand side is null for  $n$  large enough. At the point  $\sigma + \epsilon$ , the function  $\alpha(\cdot; \sigma + \epsilon)$  is not positive definite, and then there exists  $n_0$  in  $\mathbb{N}$ ,  $(\zeta_1, \dots, \zeta_{n_0})$  in  $\mathbb{R}^{n_0}$ , and  $u^0 = (u_1^0, \dots, u_{n_0}^0)$  in  $\mathbb{C}^{n_0}$  with  $\|u^0\| = 1$  such that

$$\sum_{1 \leq k, j \leq n_0} u_k^0 \alpha(\zeta_k - \zeta_j; \sigma + \epsilon) \bar{u}_j^0 < 0.$$

But with the choice (19) for the parameters, for  $n$  large enough, we obtain that for all  $k$  in  $\{1; \dots, n_0\}$ , the point  $\zeta_k$  satisfies  $|\zeta_k| \leq \frac{\ell_n}{k_n}$ , and there exists some  $\phi_n(k)$  with

$$|\zeta_k - t_{\phi_n(k),n}| \leq \frac{1}{k_n}.$$

Moreover,

$$\begin{aligned} \frac{\partial}{\partial \zeta} \alpha(\zeta; \tau) &= e^{\zeta^2 \tau^2 / 2} \left( \zeta \tau^2 \int e^{i\zeta u} h(u) du + \int i u e^{i\zeta u} h(u) du \right) \\ \left| \frac{\partial}{\partial \zeta} \alpha(\zeta; \tau) \right| &\leq e^{\zeta^2 \tau^2 / 2} (|\zeta| \Sigma^2 + \mathbb{E}_{\sigma,g} |Y_1|) \end{aligned}$$

and then for all  $\zeta, \zeta'$  belonging to the interval  $[-\ell_n/k_n; \ell_n/k_n]$ , we get

$$|\alpha(\zeta; \tau) - \alpha(\zeta'; \tau)| \leq |\zeta - \zeta'| e^{2\ell_n^2 \Sigma^2 / k_n^2} \left( \frac{\ell_n}{k_n} \Sigma^2 + \mathbb{E}_{\sigma,g} |Y_1| \right)$$

and then

$$\begin{aligned} |\alpha(\zeta_k - \zeta_j; \tau) - \alpha(t_{\phi_n(k),n} - t_{\phi_n(j),n}; \tau)| &\leq \frac{2}{k_n} e^{2\ell_n^2 \Sigma^2 / k_n^2} \left( \frac{\ell_n}{k_n} \Sigma^2 + \mathbb{E}_{\sigma,g} |Y_1| \right) \\ \left| \sum_{1 \leq k, j \leq n_0} u_k^0 [\alpha(\zeta_k - \zeta_j; \tau) - \alpha(t_{\phi_n(k),n} - t_{\phi_n(j),n}; \tau)] \bar{u}_j^0 \right| &\leq C n_0 \frac{n^a \sqrt{\log n}}{k_n} \end{aligned}$$



where  $C$  is a constant. We choose

$$k_n = n^{1/2-a-b} \text{ where } \left(\frac{1}{2} - a - b\right) > a; \text{ i.e. } 2a + b < \frac{1}{2}$$

(note that this implies in fact that  $a < 1/4$ ). Then we get

$$\sup_{\tau \in ]0; \Sigma]} \left| \sum_{1 \leq k, j \leq n_0} u_k^0 [\alpha(\zeta_k - \zeta_j; \tau) - \alpha(t_{\phi_n(k),n} - t_{\phi_n(j),n}; \tau)] \bar{u}_j^0 \right| \xrightarrow{n \rightarrow \infty} 0.$$

So we can conclude that the first term in (27) is bounded in the following way, for large enough  $n$

$$\mathbb{P}_{\sigma, g} \left( \inf_{u: \|u\|=1} \sum_{-\ell_n \leq k, l \leq \ell_n} u_k \alpha(t_{k,n} - t_{l,n}; \sigma + \epsilon) \bar{u}_l \geq -2\epsilon_n \right) \leq \mathbb{P}_{\sigma, g} \left( \sum_{-\ell_n \leq k, l \leq \ell_n} v_k \alpha(t_{k,n} - t_{l,n}; \sigma + \epsilon) \bar{v}_l \geq -2\epsilon_n \right)$$

where

$$v_k = \begin{cases} u_j^0 & \text{when } k = \phi_n(j) \text{ and } 1 \leq j \leq n_0, \\ 0 & \text{otherwise} \end{cases}$$

and  $\|v\|^2 = \sum_{k=-\ell_n}^{\ell_n} v_k^2 = \sum_{j=0}^{n_0} (u_j^0)^2 = 1$ . So we have

$$\begin{aligned} \mathbb{P}_{\sigma, g} \left( \inf_{u: \|u\|=1} \sum_{-\ell_n \leq k, l \leq \ell_n} u_k \alpha(t_{k,n} - t_{l,n}; \sigma + \epsilon) \bar{u}_l \geq -2\epsilon_n \right) &\leq \mathbb{P}_{\sigma, g} \left( \sum_{1 \leq k, l \leq n_0} u_k^0 \alpha(\zeta_k - \zeta_l; \sigma + \epsilon) \bar{u}_l^0 \right. \\ &\quad \left. \geq -2\epsilon_n + \sum_{1 \leq k, l \leq n_0} u_k^0 [\alpha(\zeta_k - \zeta_l; \sigma + \epsilon) - \alpha(t_{\phi_n(k),n} - t_{\phi_n(l),n}; \sigma + \epsilon)] \bar{u}_l^0 \right). \end{aligned}$$

We are interested in what occurs in the probability appearing in the right hand side of this inequality. All the quantities are deterministic. The first one is negative and the second one converges to zero: for  $n$  large enough, the probability of this event is null. So, for  $n$  large enough (depending on  $\epsilon$ ), we have

$$\mathbb{P}_{\sigma, g}(\hat{\sigma}_n - \sigma \geq \epsilon) \leq \Delta_n \xrightarrow{n \rightarrow \infty} 0.$$

We are studying now the probability that  $\hat{\sigma}_n$  would be less than or equal to  $\sigma - \epsilon$  for some arbitrary  $\epsilon > 0$ .

$$\begin{aligned} & \mathbb{P}_{\sigma,g}(\hat{\sigma}_n - \sigma \leq -\epsilon) \\ & \leq \mathbb{P}_{\sigma,g} \left( \exists \sigma_n \leq \sigma - \epsilon, \exists u \in \mathbb{C}^{2\ell_n}, \|u\| = 1, \sum_{-\ell_n \leq k,l \leq \ell_n} u_k \hat{\alpha}_n(t_{k,n} - t_{l,n}; \sigma_n) \bar{u}_l < -\epsilon_n \right) \\ & \leq \mathbb{P}_{\sigma,g} \left( \exists \sigma_n \leq \sigma - \epsilon, \exists u \in \mathbb{C}^{2\ell_n}, \|u\| = 1, \sum_{-\ell_n \leq k,l \leq \ell_n} u_k \alpha(t_{k,n} - t_{l,n}; \sigma_n) \bar{u}_l < 0 \right) \\ & \quad + \mathbb{P}_{\sigma,g} \left( \sup_{\tau \in [0; \Sigma]} \sup_{u; \|u\|=1} \left| \sum_{-\ell_n \leq k,l \leq \ell_n} u_k (\hat{\alpha}_n(t_{k,n} - t_{l,n}; \tau) - \alpha(t_{k,n} - t_{l,n}; \tau)) \bar{u}_l \right| > \epsilon_n \right). \end{aligned}$$

The first term in the right hand side is equal to zero as  $\alpha(\cdot, \sigma_n)$  is positive definite for  $\sigma_n$  less or equal to  $\sigma - \epsilon$ . The second one,  $\Delta_n$ , as we have already seen, tends to zero as  $n$  tends to infinity. In conclusion

$$\forall \epsilon > 0, \exists N \in \mathbb{N}, \forall n \geq N, \quad \mathbb{P}_{\sigma,g}(|\hat{\sigma}_n - \sigma| \geq \epsilon) \leq 2\Delta_n \xrightarrow[n \rightarrow \infty]{} 0. \tag{28}$$

We proved the convergence in probability of our estimator to the true value of the parameter  $\sigma$ . Note that the convergence in the  $L^2$ -norm is a consequence of the fact that  $\hat{\sigma}_n$  is almost surely bounded.  $\square$

*Proof of Proposition 3.2.* We denote by  $\Psi_g$  the Laplace transform of  $g$ :  $\Psi_g(t) = \int e^{tx} g(x) dx$ . The difference between  $\sigma$  and its estimator writes

$$\begin{aligned} \hat{\sigma}_{L,n}^2 - \sigma^2 &= \frac{2}{t_n^2} \log \left( \frac{1}{n} \sum_{j=1}^n e^{t_n Y_j - \sigma^2 t_n^2 / 2} \right) \\ &= \frac{2}{t_n^2} \log \left[ 1 + \frac{1}{n} \sum_{j=1}^n \left( \frac{e^{t_n X_j}}{\Psi_g(t_n)} e^{t_n \varepsilon_j - \sigma^2 t_n^2 / 2} - 1 \right) \right] + \frac{2}{t_n^2} \log \Psi_g(t_n) \\ &= \frac{2}{t_n^2} \log(1 + S_n) + \frac{2}{t_n^2} \log \Psi_g(t_n) \end{aligned} \tag{29}$$

where  $S_n$  is the empirical mean of independent, identically distributed centered random variables:

$$S_n \equiv \frac{1}{n} \sum_{j=1}^n \left( e^{t_n X_j} \Psi_g(t_n)^{-1} e^{t_n \varepsilon_j - \sigma^2 t_n^2 / 2} - 1 \right) \equiv \frac{1}{n} \sum_{j=1}^n U_{j,n}. \tag{30}$$

In (29), the second term in the right hand side is deterministic and converges to zero as  $n$  tends to infinity (see the definition (10) of the set  $\mathcal{L}_v$ ). Moreover

$$\text{Var}(S_n) = \frac{1}{n} \left[ \mathbb{E} \left( \frac{e^{2t_n X_1}}{\Psi_g(t_n)^2} e^{2t_n \varepsilon_1 - \sigma^2 t_n^2} \right) - 1 \right] \leq \frac{1}{n} \times \frac{\Psi_g(2t_n)}{\Psi_g(t_n)^2} e^{\sigma^2 t_n^2}.$$

Using the definition of  $\mathcal{L}_v$  (see (10)), we obtain that for all integer  $n$ :

$$e^{-t_n^2 v(t_n)} \leq \Psi_g(t_n) \leq e^{t_n^2 v(t_n)}$$

and then

$$\text{Var}(S_n) \leq \frac{1}{n} e^{(\sigma^2 + 2v(t_n) + 4v(2t_n))t_n^2} \leq \frac{e^{2t_n^2 \Sigma^2}}{n}$$

for sufficiently large  $n$  (depending on  $v$ ), and all  $\sigma \leq \Sigma$ .

With our choice of the parameters  $(t_n)_{n \geq 0}$ , we get that  $S_n$  converges in the  $L^2$ -norm, and also in probability (uniformly with respect to  $\sigma \in [0; \Sigma]$  and to  $g$  in  $\mathcal{L}_v$ ). Then equality (29) leads to the convergence of  $\hat{\sigma}_{L,n}^2$  to  $\sigma^2$  in  $\mathbb{P}_{(\sigma,g)}$ -probability, uniformly in  $\sigma$  in  $[0; \Sigma]$  and in  $g$  in  $\mathcal{L}_{v,M}$ .

Now we compute its rate of convergence with respect to the  $L^2$ -norm. We first restrict our attention to the behaviour of the term  $\log(1 + S_n)$  appearing in (29). Note that

$$\mathbb{E}[\log(1 + S_n)]^2 = \mathbb{E} \left[ \int_0^1 \frac{S_n}{1 + tS_n} dt \right]^2 \leq \int_0^1 \mathbb{E} \left( \frac{S_n^2}{(1 + tS_n)^2} \right) dt$$

Using the convexity of the function  $x \mapsto 1/x^2$  on  $\mathbb{R}^{+*}$ , we get:

$$\frac{1}{(1 + tS_n)^2} = \left[ \frac{1}{n} \sum_{j=1}^n (1 + tU_{j,n}) \right]^{-2} \leq \frac{1}{n} \sum_{j=1}^n \frac{1}{(1 + tU_{j,n})^2},$$

so that:

$$\mathbb{E} \left( \frac{S_n^2}{(1 + tS_n)^2} \right) \leq \mathbb{E} \left( \frac{S_n^2}{(1 + tU_{1,n})^2} \right) = \frac{1}{n^2} \sum_{1 \leq k, l \leq n} \mathbb{E} \left( \frac{U_{k,n} U_{l,n}}{(1 + tU_{1,n})^2} \right).$$

Now, if  $k$  differs from  $l$  and is greater or equal to 2, we have:

$$\mathbb{E} \left( \frac{U_{k,n} U_{l,n}}{(1 + tU_{1,n})^2} \right) = \mathbb{E}(U_{k,n}) \mathbb{E} \left( \frac{U_{l,n}}{(1 + tU_{1,n})^2} \right) = 0.$$

So that the sum reduces to:

$$\sum_{1 \leq k, l \leq n} \mathbb{E} \left( \frac{U_{k,n} U_{l,n}}{(1 + tU_{1,n})^2} \right) = \mathbb{E} \left( \frac{U_{1,n}^2}{(1 + tU_{1,n})^2} \right) + (n - 1) \mathbb{E}(U_{1,n}^2) \mathbb{E} \left( \frac{1}{(1 + tU_{1,n})^2} \right).$$

Returning to the upper bound on  $\mathbb{E}[\log(1 + S_n)]^2$ , we get:

$$\begin{aligned} \mathbb{E}[\log(1 + S_n)]^2 &\leq \int_0^1 \left[ \frac{1}{n^2} \mathbb{E} \left( \frac{U_{1,n}^2}{(1 + tU_{1,n})^2} \right) + \frac{1}{n} \left( 1 - \frac{1}{n} \right) \mathbb{E}(U_{1,n}^2) \mathbb{E} \left( \frac{1}{(1 + tU_{1,n})^2} \right) \right] dt \\ &\leq \frac{1}{n^2} \int_0^1 \mathbb{E} \left( \frac{U_{1,n}^2}{(1 + tU_{1,n})^2} \right) dt + \frac{1}{n} \left( 1 - \frac{1}{n} \right) \mathbb{E}(U_{1,n}^2) \int_0^1 \mathbb{E} \left( \frac{1}{(1 + tU_{1,n})^2} \right) dt. \end{aligned} \tag{31}$$

Remember the definition of the random variable  $U_{1,n}$  (see (30)) to get that:

$$\mathbb{E}(U_{1,n}^2) \leq \Psi_g(2t_n) \Psi_g(t_n)^{-2} \leq e^{2t_n^2 \Sigma^2}. \tag{32}$$

Moreover

$$\begin{aligned}
\int_0^1 \mathbb{E} \left( \frac{1}{(1+tU_{1,n})^2} \right) dt &= \mathbb{E} \int_0^1 \frac{dt}{1+tU_{1,n}} = \mathbb{E} \left[ \frac{-1}{U_{1,n}(1+tU_{1,n})} \right]_{t=0}^1 \\
&= \mathbb{E} \left( \frac{1}{U_{1,n}} \right) - \mathbb{E} \left( \frac{1}{U_{1,n}(1+U_{1,n})} \right) = \mathbb{E} \left( \frac{1}{1+U_{1,n}} \right) \\
&= \mathbb{E} \left( \Psi_g(t_n) e^{-t_n(X_1+\varepsilon_1)+\sigma^2 t_n^2/2} \right) = \Psi_g(t_n) \Psi_g(-t_n) e^{\sigma^2 t_n^2} \\
&\leq e^{2t_n^2 \Sigma^2},
\end{aligned} \tag{33}$$

for sufficiently large  $n$  (depending on  $v$ ), and all  $\sigma \leq \Sigma$ . And at least:

$$\begin{aligned}
\int_0^1 \mathbb{E} \left( \frac{U_{1,n}^2}{(1+tU_{1,n})^2} \right) dt &= \mathbb{E} \left[ \frac{-U_{1,n}}{(1+tU_{1,n})} \right]_{t=0}^1 = \mathbb{E}(U_{1,n}) - \mathbb{E} \left( \frac{U_{1,n}}{1+U_{1,n}} \right) \\
&= -\mathbb{E} \left( \frac{U_{1,n}}{1+U_{1,n}} \right) = \mathbb{E} \left[ \Psi_g(t_n) e^{-t_n(X_1+\varepsilon_1)+\sigma^2 t_n^2/2} \left( 1 - e^{t_n X_1 + t_n \varepsilon_1 + \sigma^2 t_n^2/2} \Psi_g(t_n)^{-1} \right) \right] \\
&= \Psi_g(t_n) \Psi_g(-t_n) e^{\sigma^2 t_n^2} - e^{\sigma^2 t_n^2} \\
&\leq e^{2t_n^2 \Sigma^2},
\end{aligned} \tag{34}$$

for sufficiently large  $n$  (depending on  $v$ ), and all  $\sigma \leq \Sigma$ . Combining the upper bounds (32, 33) and (34) with the inequality (31), we get that for sufficiently large  $n$  (depending on  $v$ ), and all  $\sigma \leq \Sigma$ :

$$\mathbb{E}[\log(1+S_n)]^2 \leq \frac{e^{2t_n^2 \Sigma^2}}{n^2} + \frac{1}{n} \left( 1 - \frac{1}{n} \right) e^{4t_n^2 \Sigma^2} \leq \frac{e^{4t_n^2 \Sigma^2}}{n} \leq \frac{1}{n^{1-\alpha}}.$$

Returning to the equality (29) and using a Cauchy–Schwarz inequality combined with the preceding result, we get:

$$\begin{aligned}
\sup_{\sigma \in [0; \Sigma]} \sup_{g \in \mathcal{L}_v} [\mathbb{E}(\hat{\sigma}_{L,n} - \sigma^2)^2]^{1/2} &\leq \frac{2}{t_n^2} [\mathbb{E}(\log(1+S_n))^2]^{1/2} + 2v(t_n) \\
&\leq \frac{8\Sigma^2}{\alpha(\log n)n^{(1-\alpha)/2}} + 2v(t_n),
\end{aligned}$$

which achieves the proof.  $\square$

*Proof of Theorem 4.3.* In order to simplify notations, we only give the proof for  $\ell = 1$  and  $P : x \mapsto x$ , the same arguments apply for greater values of  $\ell$ . Note first that the density  $h$  of the distribution of the observations  $Y_n$  is the convolution product between  $g$  and  $\Phi_\sigma$ . Differentiating this expression, we get that

$$\Gamma_x(y_0) \triangleq \int xg(x)\Phi_\sigma(x-y_0)dx = y_0h(y_0) + \sigma^2h'(y_0).$$

Now use the same path as in the proof of Proposition 2.1 (with the conventions  $x_0 = 0$  and  $\sigma_0 = 1$ ), assuming moreover that  $g_0$  is continuously differentiable with  $\|g'_0\|_\infty < \infty$ . The linear functional of interest writes:

$$\Gamma_t(y_0) \triangleq \int xg_t(x)\Phi_{\sigma_t}(x-y_0)dx = y_0h_t(y_0) + \sigma_t^2h'_t(y_0) = y_0h_t(y_0) + (1-t)h'_t(y_0), \quad 0 \leq t \leq \tau^2.$$

The van Trees inequality gives us:

$$\begin{aligned} \inf_{\hat{\Gamma}_n} \sup_{\Gamma \in \mathcal{G}_P(g_0, \tau)} \mathbb{E}(\hat{\Gamma}_n - \Gamma(y_0))^2 &\geq \inf_{\hat{\Gamma}_n} \int_0^{\tau^2} \mathbb{E}_t(\hat{\Gamma}_n - \Gamma_t(y_0))^2 \lambda(t) dt \\ &\geq \left[ \int \frac{\partial}{\partial t} \Gamma_t(y_0) \lambda(t) dt \right]^2 \left( n \mathbb{E}_\lambda(I) + \frac{J_0}{\tau^4} \right)^{-1}. \end{aligned}$$

We already computed the denominator of this expression in the proof of Proposition 2.1 and the only thing to do is to compute the numerator. We have:

$$\frac{\partial}{\partial t} \Gamma_t(y_0) = y_0 \frac{\partial}{\partial t} h_t(y_0) - h'_t(y_0) + (1-t) \frac{\partial}{\partial t} h'_t(y_0).$$

By the continuity of  $t \mapsto (\partial h_t)/(\partial t)(y_0)$  at the point  $t = 0$  (see the proof of Lem. 5.1), we get

$$\frac{\partial}{\partial t} h_t(y_0) = \frac{C_\tau e^{-1/2\tau^2}}{2\pi \tau} f_0(y_0)(1 + o(1)) = O(\tau).$$

The same argument applies to prove that

$$\frac{\partial}{\partial t} h'_t(y_0) = \frac{C_\tau e^{-1/2\tau^2}}{2\pi \tau} \tilde{f}_0(y_0)(1 + o(1)) = O(\tau),$$

where  $\tilde{f}_0(y_0) = \int [1 - (v - y_0)^2] e^{-(y-v)^2/2} g'_0(v) dv$ . Using the continuity of  $t \mapsto h_t(y_0)$  at  $t = 0$ , we have:

$$h'_t(y_0) = -g'_0 * \Phi_1(y_0) + O(\tau).$$

This enables us to write:

$$\int \frac{\partial}{\partial t} \Gamma_t(y_0) \lambda(t) dt = -g'_0 * \Phi_1(y_0) + O(\tau).$$

Return to the van Trees inequality to get that:

$$\begin{aligned} \inf_{\hat{\Gamma}_n} \sup_{\Gamma \in \mathcal{G}_P(g_0, \tau)} \mathbb{E}(\hat{\Gamma}_n - \Gamma(y_0))^2 &\geq (-g'_0 * \Phi_1(y_0) + O(\tau))^2 \left( n \mathbb{E}_\lambda(I) + \frac{J_0}{\tau^4} \right)^{-1} \\ &\geq (-g'_0 * \Phi_1(y_0) + O(\tau))^2 \left[ n \frac{C_\tau^2}{4\pi^2} \left( \int f_0(y)^2 h_0(y) dy \right) \tau^{-2} e^{-1/\tau^2} (1 + o(1)) + \frac{J_0}{\tau^4} \right]^{-1}. \end{aligned}$$

Choose  $\tau^{-1} = \sqrt{\log n}$  to obtain the desired result. □

The author is grateful to Elisabeth Gassiat for fruitful discussions and helpful advices.

### REFERENCES

- [1] R.J. Carroll and P. Hall, Optimal rates of convergence for deconvolving a density. *J. Amer. Statist. Assoc.* **83** (1988) 1184-1186.
- [2] L. Devroye, Consistent deconvolution in density estimation. *Canad. J. Statist.* **17** (1989) 235-239.
- [3] J. Fan, Asymptotic normality for deconvolution kernel density estimators. *Sankhyā Ser. A* **53** (1991) 97-110.
- [4] J. Fan, Global behavior of deconvolution kernel estimates. *Statist. Sinica* **1** (1991) 541-551.
- [5] J. Fan, On the optimal rates of convergence for nonparametric deconvolution problems. *Ann. Statist.* **19** (1991) 1257-1272.

- [6] J. Fan, Adaptively local one-dimensional subproblems with application to a deconvolution problem. *Ann. Statist.* **21** (1993) 600-610.
- [7] W. Feller, *An introduction to probability theory and its applications*, Vol. II. John Wiley & Sons Inc., New York (1971).
- [8] R.D. Gill and B.Y. Levit, Applications of the Van Trees inequality: A Bayesian Cramér-Rao bound. *Bernoulli* **1** (1995) 59-79.
- [9] H. Ishwaran, Information in semiparametric mixtures of exponential families. *Ann. Statist.* **27** (1999) 159-177.
- [10] B.G. Lindsay, Exponential family mixture models (with least-squares estimators). *Ann. Statist.* **14** (1986) 124-137.
- [11] M.C. Liu and R.L. Taylor, A consistent nonparametric density estimator for the deconvolution problem. *Canad. J. Statist.* **17** (1989) 427-438.
- [12] C. Matias and M.-L. Taupin, *Minimax estimation of some linear functionals in the convolution model*, Manuscript. Université Paris-Sud (2001).
- [13] P. Medgyessy, Decomposition of superposition of density functions on discrete distributions. II. *Magyar Tud. Akad. Mat. Fiz. Oszk. Közl.* **21** (1973) 261-382.
- [14] M.H. Neumann, On the effect of estimating the error density in nonparametric deconvolution. *J. Nonparametr. Statist.* **7** (1997) 307-330.
- [15] M. Pensky and B. Vidakovic, Adaptive wavelet estimator for nonparametric density deconvolution. *Ann. Statist.* **27** (1999) 2033-2053.
- [16] L. Stefanski and R.J. Carroll, Deconvoluting kernel density estimators. *Statistics* **21** (1990) 169-184.
- [17] L.A. Stefanski, Rates of convergence of some estimators in a class of deconvolution problems. *Statist. Probab. Lett.* **9** (1990) 229-235.
- [18] M.L. Taupin. Semi-parametric estimation in the non-linear errors-in-variables model. *Ann. Statist.* **29** (2001) 66-93.
- [19] A.W. van der Vaart, *Asymptotic statistics*. Cambridge University Press, Cambridge (1998).
- [20] A.W. van der Vaart and J.A. Wellner, *Weak convergence and empirical processes*. Springer-Verlag, New York (1996). With applications to statistics.
- [21] C.-H. Zhang, Fourier methods for estimating mixing densities and distributions. *Ann. Statist.* **18** (1990) 806-831.