Joint behaviour of semirecursive kernel estimators of the location and of the size of the mode of a probability density function

Abdelkader Mokkadem, Mariane Pelletier, Baba Thiam

Joint behaviour of semirecursive kernel estimators of the location and of the size of the mode of a probability density function

Abdelkader Mokkadem Mariane Pelletier Baba Thiam

(mokkadem, pelletier, thiam)@math.uvsq.fr

Université de Versailles-Saint-Quentin
Département de Mathématiques
45, Avenue des Etats-Unis
78035 Versailles Cedex
France

Abstract: Let θ and μ denote the location and the size of the mode of a probability density. We study the joint convergence rates of semirecursive kernel estimators of θ and μ. We show how the estimation of the size of the mode allows to measure the relevance of the estimation of its location. We also enlighten that, beyond their computational advantage on nonrecursive estimators, the semirecursive estimators are preferable to use for the construction on confidence regions.

AMS Subj. Classification: 62G07, 62G20
Key words : Location and size of the mode; semirecursive estimation; central limit theorem; law of the iterated logarithm

1 Introduction

Let $X_1, \ldots, X_n$ be independent and identically distributed $\mathbb{R}^d$-valued random variables with unknown probability density f. The aim of this paper is to study the joint kernel estimation of the location θ and of the size $\mu = f(\theta)$ of the mode of f. The mode is assumed to be unique, that is, $f(x) < f(\theta)$ for any $x \neq \theta$, and nondegenerated, that is, the second order differential $D^2 f(\theta)$ at the point θ is nonsingular (in the sequel, $D^m g$ will denote the differential of order m of a multivariate function g).

The problem of estimating the location of the mode of a probability density was widely studied. Kernel methods were considered, among many others, by Parzen [18], Nadaraya [17], Van Ryzin [26], Rüschendorf [23], Konakov [10], Samanta [24], Eddy ([5], [6]), Romano [20], Tsybakov [25], Vieu [27], Mokkadem and Pelletier [13], and Abraham et al. ([1], [2]). At our knowledge, the behaviour of estimators of the size of the mode has not been investigated in detail, whereas there are at least two statistical motivations for estimating this parameter. First, a use of an estimator of the size is necessary for the construction of confidence regions for the location of the mode (see, e.g., Romano [20]). As a more important motivation, let us underline that the high of the peak gives information on the shape of a density; from this point view, as suggested by Vieu [27], the location of the mode is more related to the shape of the derivative of f, whereas the size of the mode is more related to the shape of the density itself. Moreover, the knowledge of the size of the mode allows to measure the pertinence of the parameter location of the mode.

Let us mention that, even if the problem of estimating the size of the mode was not investigated in the framework of density estimation, it was studied in the framework of regression estimation. Müller [16] proves in particular the joint asymptotic normality and independence of kernel estimators of the location and of the size of the mode in the framework of nonparametric regression models with fixed design. In the framework of nonparametric regression with random design, a similar result is obtained by Ziegler ([32], [33]) for kernel estimators, and by Mokkadem and Pelletier [14] for estimators issued from stochastic approximation methods.

This paper is focused on semirecursive kernel estimators of θ and $f(\theta)$. To explain why we chose this option of semirecursive estimators, let us first recall that the (nonrecursive) wellknown kernel estimator of the location of the mode introduced by Parzen [18] is defined as a random variable θ_n^* satisfying

$$f_n^*(\theta_n^*) = \sup_{y \in \mathbb{R}^d} f_n^*(y),$$

where f_n^* is Rosenblatt's estimator of f; more precisely,

$$f_n^*(x) = \frac{1}{n h_n^d} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h_n}\right),$$

where the bandwidth (h_n) is a sequence of positive real numbers going to zero and the kernel K is a continuous function satisfying $\lim_{\|x\| \to +\infty} K(x) = 0$, $\int_{\mathbb{R}^d} K(x)dx = 1$. The asymptotic behaviour of θ_n^* was widely studied (see, among others, [5], [6], [10], [13], [17], [18], [20], [23], [24], [26], [27]), but, on a computational point of view, the estimator θ_n^* has a main drawback: its update, from a sample size n to a sample size $n+1$, is far from being immediate. Applying the stochastic approximation method, Tsybakov [25] introduced the recursive kernel estimator of θ defined as

$$T_n = T_{n-1} + \gamma_n \frac{1}{h_n^{d+1}} \nabla K\left(\frac{T_{n-1} - X_n}{h_n}\right),$$

where $T_0 \in \mathbb{R}^d$ is arbitrarily chosen, and the stepsize (γ_n) is a sequence of positive real numbers going to zero. The great property of this estimator is that its update is very rapid. Unfortunately, for reasons inherent to stochastic approximation algorithms properties, very strong assumptions on the density f must be required to ensure its consistency. A recursive version f_n of Rosenblatt's density estimator was introduced by Wolverton and Wagner [30] (and discussed, among others, by Yamato [31], Davies [3], Devroye [4], Menon et al. [12], Wertz [29], Wegman and Davies [28], Roussas [22], and Mokkadem et al. [15]). Let us recall that f_n is defined as

$$f_n(x) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{h_i^d} K\left(\frac{x - X_i}{h_i}\right). \tag{1}$$

Its update from a sample of size n to one of size $n+1$ is immediate since f_n clearly satisfies the recursive relation

$$f_n(x) = \left(1 - \frac{1}{n}\right) f_{n-1}(x) + \frac{1}{nh_n^d} K\left(\frac{x - X_n}{h_n}\right).$$

This property of rapid update of the density estimator is particularly important in the framework of mode estimation, since the number of points where f must be estimated is very large. We thus define a semirecursive version of Parzen's estimator of the location of the mode by using Wolverton-Wagner's recursive density estimator, rather than Rosenblatt's density estimator. More precisely, our estimator θ_n of the location θ of the mode is a random variable satisfying

$$f_n(\theta_n) = \sup_{y \in \mathbb{R}^d} f_n(y). \tag{2}$$

Let us mention that, in the same way as for Parzen's estimator, the fact that the kernel K is continuous and vanishing at infinity ensures that the choice of θ_n as a random variable satisfying (2) can be made with the help of an order on $\mathbb{R}^d$. For example, one can consider the following lexicographic order: $x \leq y$ if the first nonzero coordinate of $x - y$ is negative. The definition

$$\theta_n = \inf\left\{ y \in \mathbb{R}^d \text{ such that } f_n(y) = \sup_{x \in \mathbb{R}^d} f_n(x) \right\},$$

where the infimum is taken with respect to the lexicographic order on $\mathbb{R}^d$, ensures the measurability of the kernel mode estimator.

Let us also mention that, in order to make more rapid the computation of the kernel estimator of the location of the mode, Abraham et al. ([1], [2]) proposed the following alternative version of Parzen's estimator θ_n^*:

$$\hat{\theta}_n^* = \underset{1 \leq i \leq n}{\operatorname{argmax}} f_n^*(X_i).$$

Similarly, we could consider the following alternative version of our semirecursive estimator θ_n:

$$\hat{\theta}_n = \underset{1 \leq i \leq n}{\operatorname{argmax}} f_n(X_i).$$

However, to establish the asymptotic properties of $\hat{\theta}_n^*$, Abraham et al. [2] prove the asymptotic proximity between θ_n^* and $\hat{\theta}_n^*$, which allows them to deduce the asymptotic weak behaviour of $\hat{\theta}_n^*$ from the one of θ_n^*. In the same way, we can conjecture that the asymptotic weak behaviour of $\hat{\theta}_n$ could be deduced from the one of θ_n, but, in this paper, we limit ourselves on establishing the

asymptotic properties of θ_n.

Let us now come back to the problem of estimating the size $f(\theta)$ of the mode. The ordinarily used estimator is defined as $\mu_n^* = f_n^*(\theta_n^*)$ (f_n^* being Rosenblatt's density estimator and θ_n^* Parzen's mode estimator); the consistency of μ_n^* is sufficient to allow the construction of confidence regions for θ (see, e.g., Romano [20]). Adapting the construction of μ_n^* to the semirecursive framework would lead us to estimate $f(\theta)$ by

$$\mu_n = f_n(\theta_n). \tag{3}$$

However, this estimator has two main drawbacks (as well as μ_n^*). First, the use of a higher order kernel K is necessary for $(\mu_n - \mu)$ to satisfy a central limit theorem, and thus for the construction of confidence intervals of μ (and of confidence regions for (θ, μ)). Moreover, in the case when a higher order kernel is used, it is not possible to choose a bandwidth for which both estimators θ_n and μ_n converge at the optimal rate. These constations lead us to use two different bandwidths, one for the estimation of θ, the other one for the estimation of μ. More precisely, let $\tilde{f}_n$ be the recursive kernel density estimator defined as

$$\tilde{f}_n(x) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{\tilde{h}_i^d} K\left(\frac{x - X_i}{\tilde{h}_i}\right),$$

where the bandwidth $(\tilde{h}_n)$ may be different from (h_n) used in the definition of f_n (see (1)); we estimate the size of the mode by

$$\tilde{\mu}_n = \tilde{f}_n(\theta_n), \tag{4}$$

where θ_n is still defined by (2), and thus with the first bandwidth (h_n).

The purpose of this paper is the study of the joint asymptotic behaviour of θ_n and $\tilde{\mu}_n$. We first prove the strong consistency of both estimators. We then establish the joint weak convergence rate of θ_n and $\tilde{\mu}_n$. We prove in particular that adequate choices of the bandwidths lead to the asymptotic normality and independence of these estimators, and that the use of different bandwidths allow to obtain simultaneously the optimal convergence rate of both estimators. We then apply our weak convergence rate result to the construction of confidence regions for (θ, μ), and illustrate this application with a simulations study. This application enlightens the advantage of using semirecursive estimators rather than nonrecursive estimators. It also shows how the estimation of the size of the mode gives information on the relevance of estimating its location. Finally, we establish the joint strong convergence rate of θ_n and $\tilde{\mu}_n$.

2 Assumptions and Main Results

Throughout this paper, (h_n) and $(\tilde{h}_n)$ are defined as $h_n = h(n)$ and $\tilde{h}_n = \tilde{h}(n)$ for all $n \geq 1$, where h and $\tilde{h}$ are two positive functions.

2.1 Strong consistency

The conditions we require for the strong consistency of θ_n and $\tilde{\mu}_n$ are the following.

(A1) i) K is an integrable, differentiable, and even function such that $\int_{\mathbb{R}^d} K(z)dz = 1$.
ii) There exists $\zeta > 0$ such that $\int_{\mathbb{R}^d} \|z\|^\zeta |K(z)dz| < \infty$.
iii) K is Hölder continuous.
iv) There exists $\gamma > 0$ such that $z \mapsto \|z\|^\gamma |K(z)|$ is a bounded function.

(A2) i) f is uniformly continuous on $\mathbb{R}^d$.
ii) There exists $\xi > 0$ such that $\int_{\mathbb{R}^d} \|x\|^\xi f(x)dx < \infty$.
iii) There exists $\eta > 0$ such that $z \mapsto \|z\|^\eta f(z)$ is a bounded function.
iv) There exists $\theta \in \mathbb{R}^d$ such that $f(x) < f(\theta)$ for all $x \neq \theta$.

(A3) The functions h and $\tilde{h}$ are locally bounded and vary regularly with exponent $(-a)$ and $(-\tilde{a})$ respectively, where $a \in]0, 1/(d+4)[$, $\tilde{a} \in]0, 1/(d+2)[$.

Remark 1 *Note that (A1)iv) implies that K is bounded.*

Remark 2 *The assumptions required on the probability density to establish the strong consistency of the semirecursive estimator of the location of the mode are slightly stronger than those needed for the nonrecursive estimator (see, e.g., [13], [20]), but are much weaker than the ones needed for the recursive estimator (see [25]).*

Remark 3 *Let us recall that a positive function (not necessarily monotone) $\mathcal{L}$ defined on $]0, \infty[$ is slowly varying if $\lim_{t\to\infty} \mathcal{L}(tx)/\mathcal{L}(t) = 1$, and that a function G varies regularly with exponent ρ, $\rho \in \mathbb{R}$, if and only if it is of the form $G(x) = x^\rho \mathcal{L}(x)$ with $\mathcal{L}$ slowly varying (see, for example, Feller [8] page 275). Typical examples of regularly varying functions are x^ρ, $x^\rho \log x$, $x^\rho \log\log x$, $x^\rho \log x/\log\log x$, and so on.*

Proposition 1 *Let θ_n and $\tilde{\mu}_n$ be defined by (2) and (4), respectively. Under (A1)-(A3),*

$$\lim_{n\to\infty} \theta_n = \theta \quad a.s. \quad and \quad \lim_{n\to\infty} \tilde{\mu}_n = \mu \quad a.s.$$

2.2 Weak convergence rate

In order to state the weak convergence rate of θ_n and $\tilde{\mu}_n$, we need the following additional assumptions on K and f.

(A4) i) K is twice differentiable on $\mathbb{R}^d$.
ii) $z \mapsto z\nabla K(z)$ is integrable.
iii) For any $(i,j) \in \{1,\ldots,d\}^2$, $\partial^2 K/\partial x_i \partial x_j$ is bounded integrable and Hölder continuous.
iv) K is a kernel of order $q \geq 2$ i.e. $\forall s \in \{1,\ldots,q-1\}$, $\forall j \in \{1,\ldots,d\}$, $\int_{\mathbb{R}^d} y_j^s K(y)dy_j = 0$ and $\int_{\mathbb{R}^d} |y_j^q K(y)|dy < \infty$.

(A5) i) $D^2 f(\theta)$ is nonsingular.
ii) $D^2 f$ is q-times differentiable, ∇f and $D^q f$ are bounded.
iii) For any $(i,j) \in \{1,\ldots,d\}^2$, $\sup_{x\in\mathbb{R}^d} \|D^q\left(\partial^2 f/\partial x_i \partial x_j\right)\| < \infty$, and for any $k \in \{1,\ldots,d\}$, $\sup_{x\in\mathbb{R}^d} \|D^q(\partial f/\partial x_k)\| < \infty$.

Remark 4 *Note that (A4)ii) and (A4)iii) imply that ∇K is Lipschitz-continuous and integrable; it is thus straightforward to see that $\lim_{\|x\|\to\infty} \|\nabla K(x)\| = 0$ (and in particular ∇K is bounded).*

We also need to add conditions on the bandwiths. Let us set

$$\mathcal{L}_\theta(n) = n^a h_n \quad \text{and} \quad \mathcal{L}_\mu(n) = n^{\tilde{a}} \tilde{h}_n.$$

(In view of (A3), $\mathcal{L}_\theta$ and $\mathcal{L}_\mu$ are positive slowly varying functions, see Remark 3). In the statement of the the weak convergence rate of θ_n and $\tilde{\mu}_n$, we shall refer to the following conditions.

(C1) One of the following two conditions is fulfilled.

i) $\dfrac{1}{d+4} < \tilde{a} < \dfrac{q}{d+2q+2}$ and $\dfrac{\tilde{a}}{q} < a < \dfrac{1-2\tilde{a}}{d+2}$;

ii) $\dfrac{1}{d+2q} < \tilde{a} \leq \dfrac{1}{d+4}$ and $\dfrac{1}{d+2q+2} < a < \dfrac{1+\tilde{a}d}{2\,(d+2)}$

(C2) One of the following two conditions is fulfilled.

i) $0 < \tilde{a} < \dfrac{1}{d+2q}$ and $\dfrac{\tilde{a}}{2} < a < \dfrac{1}{d+2q+2}$;

ii) $\tilde{a} = \dfrac{1}{d+2q}$, $\lim_{n\to\infty} \mathcal{L}_\mu(n) = \infty$ and $\dfrac{1}{2(d+2q)} < a < \dfrac{1}{d+2q+2}$.

Remark 5 *(C1) implies that* $\lim_{n\to\infty} nh_n^{d+2q+2} = 0$ *and* $\lim_{n\to\infty} n\tilde{h}_n^{d+2q} = 0$*, whereas (C2) implies that* $\lim_{n\to\infty} nh_n^{d+2q+2} = \infty$ *and* $\lim_{n\to\infty} n\tilde{h}_n^{d+2q} = \infty$*.*

We finally need to introduce the following notation:

$$B_q(\theta) = \begin{pmatrix} \frac{(-1)^q}{q!(1-aq)} \nabla\Big(\sum_{j=1}^d \beta_j^q \frac{\partial^q f}{\partial x_j^q}(\theta) \Big) \\ \frac{(-1)^q}{q!(1-\tilde{a}q)} \sum_{j=1}^d \beta_j^q \frac{\partial^q f}{\partial x_j^q}(\theta) \end{pmatrix} \quad \text{with} \quad \beta_j^q = \int_{\mathbb{R}^d} y_j^q K(y)dy, \quad aq \neq 1 \ \text{ and } \ \tilde{a}q \neq 1, \tag{5}$$

$$A = \begin{pmatrix} -\left[D^2 f(\theta)\right]^{-1} & 0 \\ 0 & 1 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \frac{f(\theta) G}{1+a(d+2)} & 0 \\ 0 & \frac{f(\theta)\int_{\mathbb{R}^d} K^2(z)dz}{1+\tilde{a}d} \end{pmatrix}, \tag{6}$$

G is the matrix $d \times d$ defined by $G^{(i,j)} = \displaystyle\int_{\mathbb{R}^d} \frac{\partial K}{\partial x_i}(x) \frac{\partial K}{\partial x_j}(x)dx$, and, for any $c, \tilde{c} \geq 0$, $D(c,\tilde{c}) = \begin{pmatrix} \sqrt{c} I_d & 0 \\ 0 & \sqrt{\tilde{c}} \end{pmatrix}$ where I_d is the $d \times d$ identity matrix.

Theorem 1 *Let* θ_n *and* $\tilde{\mu}_n$ *be defined by (2) and (4), respectively, and assume that (A1)-(A5) hold.*

i) If (C1) is satisfied, then

$$\begin{pmatrix} \sqrt{nh_n^{d+2}}(\theta_n - \theta) \\ \sqrt{n\tilde{h}_n^d}(\tilde{\mu}_n - \mu) \end{pmatrix} \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, A\Sigma A\right).$$

ii) If $a = (d+2q+2)^{-1}$, $\tilde{a} = (d+2q)^{-1}$*, and if there exist* $c, \tilde{c} \geq 0$ *such that* $\lim_{n\to\infty} nh_n^{d+2q+2} = c$ *and* $\lim_{n\to\infty} n\tilde{h}_n^{d+2q} = \tilde{c}$*, then*

$$\begin{pmatrix} \sqrt{nh_n^{d+2}}(\theta_n - \theta) \\ \sqrt{n\tilde{h}_n^d}(\tilde{\mu}_n - \mu) \end{pmatrix} \xrightarrow{\mathcal{D}} \mathcal{N}\left(D(c,\tilde{c}) A B_q(\theta), A\Sigma A\right).$$

iii) If (C2) is satisfied, then

$$\begin{pmatrix} \frac{1}{h_n^q}(\theta_n - \theta) \\ \frac{1}{\tilde{h}_n^q}(\tilde{\mu}_n - \mu) \end{pmatrix} \xrightarrow{\mathbb{P}} A B_q(\theta).$$

Remark 6 *The simultaneous weak convergence rate of nonrecursive estimators of the location and size of the mode can be established by following the lines of the proof of Theorem 1. More precisely, set*

$$B_q^*(\theta) = \begin{pmatrix} \frac{(-1)^q}{q!}\nabla\Big(\sum_{j=1}^d \beta_j^q \frac{\partial^q f}{\partial x_j^q}(\theta)\Big) \\ \frac{(-1)^q}{q!}\sum_{j=1}^d \beta_j^q \frac{\partial^q f}{\partial x_j^q}(\theta) \end{pmatrix}, \ \Sigma^* = \begin{pmatrix} f(\theta)G & 0 \\ 0 & f(\theta)\int_{\mathbb{R}^d} K^2(z)dz \end{pmatrix},$$

let θ_n^* be Parzen's kernel estimator of the location of the mode and $\tilde{\mu}_n^* = \tilde{f}_n^*(\theta_n^*)$ be the kernel estimator of the size of the mode defined with the help of θ_n^* and of Rosenblatt's density estimator $\tilde{f}_n^*$ (the bandwidth $(\tilde{h}_n)$ defining $\tilde{f}_n^*$ being eventually different from the banwidth (h_n) used to define θ_n^*); Theorem 1 holds when θ_n, $\tilde{\mu}_n$, $B_q(\theta)$, Σ are replaced by θ_n^*, $\tilde{\mu}_n^*$, $B_q^*(\theta)$, Σ^*, respectively.

Part 1 and Part 2 in the case $c = \tilde{c} = 0$ (respectively Part 3) of Theorem 1 correspond to the case when the bias (respectively the variances) of both estimators θ_n and $\tilde{\mu}_n$ are negligeable in front of their respective variances (respectively bias). When $c, \tilde{c} > 0$, Part 2 of Theorem 1 corresponds to the case when the bias and the variance of each estimator θ_n and $\tilde{\mu}_n$ have the same convergence rate. Other possible conditions lead to different combinations; these ones have been omitted for sake of simplicity.

Theorem 1 gives the joint weak convergence rate of θ_n and $\tilde{\mu}_n$. Of course, it is also possible to estimate the location and the size of the mode separately. Concerning the estimation of the location of the mode, let us enlighten that the advantage of the semirecursive estimator θ_n on its nonrecursive version θ_n^* is that its asymptotic variance $[1+a(d+2)]^{-1}f(\theta)G$ is smaller than the one of Parzen's estimator, which equals $f(\theta)G$ (see, e.g. Romano [20] for the case $d = 1$ and Mokkadem and Pelletier [13] for the case $d \geq 1$); this advantage of semirecursive estimators will be discussed again in Section 2.3. The estimation of the size of the mode is of course not independent of the estimation of the location, since the estimator $\tilde{\mu}_n$ is constructed with the help of the estimator θ_n. To get a good estimation of the size of the mode, it seems obvious that θ_n should be computed with a bandwidth (h_n) leading to its optimal convergence rate (or, at least, to a convergence rate close to the optimal one). The main information given by Theorem 1 is that, for $\tilde{\mu}_n$ to converge at the optimal rate, the use of a second bandwidth $(\tilde{h}_n)$ is then necessary.

Let us enlighten that, in the case when θ_n and $\tilde{\mu}_n$ satisfy a central limit theorem (Parts 1 and 2 of Theorem 1), these estimators are asymptotically independent, although, in its definition, the estimator of the size of the mode is heavily connected to the one of the location of the mode. As pointed out by a referee, this property was expected. As a matter of fact (and as mentioned in the introduction), the location of the mode is a parameter which gives information on the shape of the density derivative, whereas the size of the mode gives information on the shape of the density itself. This constatation must be related to the fact that the weak (and strong) convergence rate of θ_n is given by the one of the gradient of f_n, whereas the weak (and strong) convergence rate of $\tilde{\mu}_n$ is given by the one of $\tilde{f}_n$ itself; the variance of the density estimators converging to zero faster than the one of the estimators of the density derivatives, the asymptotic independence of θ_n and $\tilde{\mu}_n$ is completely explained.

Let us finally say one word on our assumptions on the bandwidths. In the framework of nonrecursive estimation, there is no need to assume that (h_n) and $(\tilde{h}_n)$ are regularly varying sequences. In the case of semirecursive estimation, this assumption can obviously not be omitted, since the exponents a and $\tilde{a}$ stand in the expressions of the asymptotic bias $B_q(\theta)$ and variance Σ. This might be seen as a slight inconvenient of semirecursive estimation; however, as it is

enlightened in the following section, it turns out to be an advantage, since the asymptotic variances of the semirecursive estimators are smaller than the ones of the nonrecursive estimators.

2.3 Construction of confidence regions and simulations studies

The application of Theorem 1 (and of Remark 6) allows the construction of confidence regions (simultaneous or not) of the location and of the size of the mode, as well as confidence ellipsoids of the couple (θ, μ). Hall [9] shows that, in order to construct confidence regions, avoiding bias estimation by a slight undersmoothing is more efficient than explicit bias correction. In the framework of undersmoothing, the asymptotic bias of the estimator is negligeable in front of its asymptotic variance; according to the estimation by confidence regions point of view, the parameter to minimize is thus the asymptotic variance. Now, note that

$$\Sigma = \begin{pmatrix} [1+a(d+2)]^{-1} I_d & 0 \\ 0 & [1+\tilde{a}d]^{-1} \end{pmatrix} \Sigma^*$$

(where $A\Sigma A$ (respectively $A\Sigma^* A$) is the asymptotic covariance matrix of the semirecursive estimators $(\theta_n, \tilde{\mu}_n)$ (respectively of the nonrecursive estimators $(\theta_n^*, \tilde{\mu}_n^*)$). In order to construct confidence regions for the location and/or size of the mode, it is thus much preferable to use semirecursive estimators rather than nonrecursive estimators. Simulations studies confirm this theoritical conclusion, whatever the parameter (θ, μ or (θ, μ)) for which confidence regions are contructed is. For sake of succintness, we do not give all these simulations results here, but focuse on the construction of confidence ellipsoid for (θ, μ); the aim of this example is of course to enlighten the advantage of using semirecursive estimators rather than nonrecursive estimators, but also to show how this confidence region gives informations on the shape of the density, and, consequently allows to measure the pertinence of the parameter location of the mode.

To construct confidence regions for (θ, μ), we consider the case $d = 1$. The following corollary is a straightforward consequence of Theorem 1.

Corollary 1 *Let θ_n and $\tilde{\mu}_n$ be defined by (2) and (4), respectively, and assume that (A1)-(A5) hold. Moreover, let (h_n) and $(\tilde{h}_n)$ either satisfy (C1) or be such that $\lim_{n\to\infty} nh_n^{2q+3} = 0$ and $\lim_{n\to\infty} n\tilde{h}_n^{2q+1} = 0$ with $a = (2q+3)^{-1}$ and $\tilde{a} = (2q+1)^{-1}$. We then have*

$$\frac{(1+3a)nh_n^3[f''(\theta)]^2}{f(\theta)\int_{\mathbb{R}} K'^2(x)dx}(\theta_n - \theta)^2 + \frac{(1+\tilde{a})n\tilde{h}_n}{f(\theta)\int_{\mathbb{R}} K^2(x)dx}(\tilde{\mu}_n - \mu)^2 \quad \xrightarrow{\mathcal{D}} \quad \chi^2(2). \tag{7}$$

Moreover, (7) still holds when the parameters $f(\theta)$ and $f''(\theta)$ are replaced by consistent estimators.

Remark 7 *In view of Remark 6, in the case when the nonrecursive estimators θ_n^* and $\tilde{\mu}_n^*$ are used, (7) becomes*

$$\frac{nh_n^3[f''(\theta)]^2}{f(\theta)\int_{\mathbb{R}} K'^2(x)dx}(\theta_n^* - \theta)^2 + \frac{n\tilde{h}_n}{f(\theta)\int_{\mathbb{R}} K^2(x)dx}(\tilde{\mu}_n^* - \mu)^2 \quad \xrightarrow{\mathcal{D}} \quad \chi^2(2) \tag{8}$$

(and, again, this convergence still holds when the parameters $f(\theta)$ and $f''(\theta)$ are replaced by consistent estimators).

Let $\check{f}_n''$ (respectively $\check{f}_n^{*''}$) be the recursive estimator (respectively the nonrecursive Rosenblatt's estimator) of f'' computed with the help of a bandwidth $\check{h}_n$, and set

$$P_n = \frac{(1+3a)nh_n^3[\check{f}_n''(\theta_n)]^2}{\tilde{f}_n(\theta_n)\int_{\mathbb{R}} K'^2(x)dx}, \qquad Q_n = \frac{(1+\tilde{a})n\tilde{h}_n}{\tilde{f}_n(\theta_n)\int_{\mathbb{R}} K^2(x)dx},$$
$$P_n^* = \frac{nh_n^3[\check{f}_n^{*''}(\theta_n^*)]^2}{\tilde{f}_n^*(\theta_n^*)\int_{\mathbb{R}} K'^2(x)dx}, \qquad Q_n^* = \frac{n\tilde{h}_n}{\tilde{f}_n^*(\theta_n^*)\int_{\mathbb{R}} K^2(x)dx}.$$

Moreover, let c_α be such that $\mathbb{P}(Z \leq c_\alpha) = 1-\alpha$, where Z is $\chi^2(2)$-distributed; in view of Corollary 1 and Remark 7, the sets

$$\begin{aligned}\mathcal{E}_\alpha &= \left\{(\theta,\mu)/\ P_n(\theta_n-\theta)^2 + Q_n(\tilde{\mu}_n-\mu)^2 \leq c_\alpha\right\}\\ \mathcal{E}_\alpha^* &= \left\{(\theta,\mu)/\ P_n^*(\theta_n^*-\theta)^2 + Q_n^*(\tilde{\mu}_n^*-\mu)^2 \leq c_\alpha\right\}\end{aligned}$$

are confidence ellipsoids for (θ,μ) with asymptotic coverage level $1-\alpha$. Let us dwell on the fact that both confidence regions have the same asymptotic level, but the lengths of the axes of the first one (constructed with the help of the semirecursive estimators θ_n and $\tilde{\mu}_n$) are smaller than the ones of the second one (constructed with the help of the nonrecursive estimators θ_n^* and $\tilde{\mu}_n^*$).

We now present simulations results. In order to see the relationship between the shape of the confidence ellipsoids and the one of the density, the density f we consider is the density of the $\mathcal{N}(0,\sigma^2)$-distribution, the parameter σ taking the values 0.3, 0.4, 0.5, 0.7, 0.75, 1, 1.5, 2, and 2.5. We use the sample size $n = 100$ and the coverage level $1-\alpha = 95\%$ (and thus $c_\alpha = 5.99$). In each case, the number of simulations is $N = 5000$. The kernel we use is the standard Gaussian density; the bandwidths are

$$h_n = \frac{n^{-1/7}}{(\log n)}, \quad \tilde{h}_n = \frac{n^{-1/5}}{(\log n)}, \quad \check{h}_n = n^{-1/9}.$$

Table 1 below gives, for each value of σ, the empirical values of θ_n , θ_n^*, μ_n , μ_n^* (with respect to the 5000 simulations), and:

b the empirical length of the θ-axis of the confidence ellipsoid $\mathcal{E}_{5\%}$;

b^* the empirical length of the θ-axis of the confidence ellipsoid $\mathcal{E}_{5\%}^*$;

a the empirical length of the μ-axis of the confidence ellipsoid $\mathcal{E}_{5\%}$;

a^* the empirical length of the μ-axis of the confidence ellipsoid $\mathcal{E}_{5\%}^*$;

p the empirical coverage level of the confidence ellipsoid $\mathcal{E}_{5\%}$;

p^* the empirical coverage level of the confidence ellipsoid $\mathcal{E}_{5\%}^*$.

Table 1

σ	0.3	0.4	0.5	0.7	0.75	1	1.5	2	2.5
θ_n	−0.002	0.004	0.001	0.003	0.002	0.014	−0.005	−0.009	0.014
θ_n^*	0.003	0.005	0.001	0.005	−0.008	0.016	0.003	−0.020	−0.046
b	1.154	1.346	1.805	2.898	3.160	5.218	10.094	17.866	17.405
b^*	1.166	1.458	1.968	3.300	3.582	5.925	12.943	21.946	23.715
μ_n	1.335	0.989	0.782	0.564	0.522	0.401	0.263	0.196	0.155
μ_n^*	1.312	0.979	0.783	0.562	0.512	0.388	0.269	0.193	0.163
a	0.444	0.399	0.365	0.322	0.315	0.283	0.247	0.224	0.210
a^*	0.514	0.459	0.420	0.369	0.363	0.327	0.287	0.261	0.246
p	98.7%	97.8%	98.2%	98.4%	97.7%	97.8%	97.5%	97.2%	98.4%
p^*	98.6%	98.1%	98.4%	98.2%	96.8%	96.6%	96.9%	97.7%	98.2%

Confirming our theoritical results, we see that the empirical coverage levels of both confidence ellipsoids $\mathcal{E}^*_{5\%}$ and $\mathcal{E}_{5\%}$ are similar, but that the empirical areas of the ellipsoids $\mathcal{E}_{5\%}$ (constructed with the help of the semirecursive estimators) are always smaller than the ones of the the ellipsoids $\mathcal{E}^*_{5\%}$ (constructed with the help of the nonrecursive estimators).

Let us now discuss the interest of the estimation of the size of the mode and the one of the joint estimation of the location and size of the mode. Both estimations give informations on the shape of the probability density and, consequently, allow to measure the pertinence of the parameter location of the mode. Of course, the parameter θ is significant only in the case when the high of the peak is large enough; since we consider here the example of the $\mathcal{N}(0,\sigma^2)$-distribution, this corresponds to the case when σ is small enough. Estimating only the size of the mode gives a first idea of the shape of the density around the location of the mode (for instance, when the size is estimated around 0.16, it is clear that the density is very flat). Now, the shape of the confidence ellipsoids allows to get a more precise idea. As a matter of fact, for small values of σ, the length of the μ-axis is larger than the one of the θ-axis; as σ increases, the length of the μ-axis decreases, and the one of the θ-axis increases (for $\sigma = 2.5$, the length of the θ-axis is larger than 20 times the one of the μ-axis). Let us underline that these variations of the lengths of the axes are not due to bad estimations results; Table 2 below gives the values of the lengths b (respectively b^*) of the θ-axis, a (respectively a^*) of the μ-axis of the ellipsoids computed with the semirecursive estimators θ_n and $\tilde{\mu}_n$ (respectively with the nonrecursive estimators θ^*_n and $\tilde{\mu}^*_n$) in the case when the true values of the parameters $f(\theta)$ and $f''(\theta)$ are used (that is, by straightforwardly applying (7) and (8)).

Table 2

σ	0.3	0.4	0.5	0.7	0.75	1	1.5	2	2.5
b	0.159	0.327	0.571	1.357	1.572	3.227	8.895	18.260	31.899
b^*	0.190	0.390	0.682	1.622	1.879	3.858	10.631	21.825	38.127
μ	1.333	0.998	0.798	0.570	0.532	0.399	0.266	0.199	0.159
a	0.465	0.403	0.360	0.303	0.294	0.255	0.208	0.180	0.161
a^*	0.509	0.441	0.395	0.332	0.322	0.279	0.228	0.197	0.176

2.4 Strong convergence rate

To establish the joint strong convergence rate of θ_n and $\tilde{\mu}_n$, we need the following additionnal assumption.

(A6) i) h and $\tilde{h}$ are differentiables, their derivatives vary regularly with exponent $(-a-1)$ and $(-\tilde{a}-1)$ respectively.
ii) There exists $n_0 \in \mathbb{N}$ such that

$$n \geq m \geq n_0 \quad \Rightarrow \quad \max\left\{\frac{mh_m^{-(d+2)}}{nh_n^{-(d+2)}}; \frac{m\tilde{h}_m^{-d}}{n\tilde{h}_n^{-d}}\right\} \quad = \quad \frac{\min\left\{mh_m^{-(d+2)}; m\tilde{h}_m^{-d}\right\}}{\min\left\{nh_n^{-(d+2)}; n\tilde{h}_n^{-d}\right\}}.$$

Remark 8 *Assumption (A6)ii) holds when $a \neq \tilde{a}$, and in the case $a = \tilde{a}$, it is satisfied when $\mathcal{L}_\theta(n) = (\mathcal{L}_\mu(n))^{\frac{d}{d+2}}$ for n large enough.*

Moreover, condition (C2) is replaced by the following one.

(C'2) Either (C2) i) is fulfilled or $\tilde{a} = \dfrac{1}{d+2q}$, $\lim_{n\to\infty} \dfrac{(\mathcal{L}_\mu(n))^{d+2q}}{2\log\log n} = \infty$, and $\dfrac{1}{2(d+2q)} < a < \dfrac{1}{d+2q+2}$.

Before stating the almost sure convergence rate of $(\theta_n^T, \tilde{\mu}_n)^T$, let us remark that Proposition 2.3 in Mokkadem and Pelletier [13] ensures that the matrix G (and thus the matrix Σ) is nonsingular.

Theorem 2 *Let θ_n and $\tilde{\mu}_n$ be defined by (2) and (4), respectively, and assume that (A1)-(A6) hold.*

i) If (C1) is fulfilled, then, with probability one, the sequence

$$\frac{1}{\sqrt{2\log\log n}}\begin{pmatrix} \sqrt{nh_n^{d+2}}(\theta_n-\theta) \\ \sqrt{n\tilde{h}_n^{d}}(\tilde{\mu}_n-\mu) \end{pmatrix}$$

is relatively compact and its limit set is the ellipsoid

$$\mathcal{E} = \left\{\nu \in \mathbb{R}^{d+1} \text{ such that } \nu^T A^{-1}\Sigma^{-1}A^{-1}\nu \leq 1\right\}.$$

ii) If $a = (d+2q+2)^{-1}$, $\tilde{a} = (d+2q)^{-1}$, and if there exist $c, \tilde{c} \geq 0$ such that $\lim_{n\to\infty} nh_n^{d+2q+2}/(2\log\log n) = c$ and $\lim_{n\to\infty} n\tilde{h}_n^{d+2q}/(2\log\log n) = \tilde{c}$, then, with probability one, the sequence

$$\frac{1}{\sqrt{2\log\log n}}\begin{pmatrix} \sqrt{nh_n^{d+2}}(\theta_n-\theta) \\ \sqrt{n\tilde{h}_n^{d}}(\tilde{\mu}_n-\mu) \end{pmatrix}$$

is relatively compact and its limit set is the ellipsoid

$$\mathcal{E} = \left\{\nu \in \mathbb{R}^{d+1} \text{ such that } \left(A^{-1}\nu - D\left(c,\tilde{c}\right)B_q(\theta)\right)^T \Sigma^{-1}\left(A^{-1}\nu - D\left(c,\tilde{c}\right)B_q(\theta)\right) \leq 1\right\}.$$

iii) If (C'2) is satisfied, then

$$\begin{pmatrix} \frac{1}{h_n^q}(\theta_n-\theta) \\ \frac{1}{\tilde{h}_n^q}(\tilde{\mu}_n-\mu) \end{pmatrix} \xrightarrow{a.s.} AB_q(\theta).$$

Remark 9 *(C'1) implies that $\lim_{n\to\infty} nh_n^{d+2q+2}/\log\log n = 0$ and $\lim_{n\to\infty} n\tilde{h}_n^{d+2q}/\log\log n = 0$, whereas (C'2) implies that $\lim_{n\to\infty} nh_n^{d+2q+2}/\log\log n = \infty$ and $\lim_{n\to\infty} n\tilde{h}_n^{d+2q}/\log\log n = \infty$.*

Laws of the iterated logarithm for Parzen's nonrecursive kernel mode estimator were established by Mokkadem and Pelletier [13]. The technics of demonstration used in the framework of nonrecursive estimators are totally different from those employed to prove Theorem 2. This is due to the following fondamental difference between the nonrecursive estimator θ_n^* and the semirecursive estimator θ_n: the study of the asymptotic behaviour of θ_n^* comes down to the one of a *triangular sum* of independent variables, whereas the study of the asymptotic behaviour of θ_n reduces to the one of a *sum* of independent variables. Of course, this difference is not quite important for the study of the weak convergence rate. But, for the study of the strong convergence rate, it makes the case of the semirecursive estimation much easier than the case of the nonrecursive estimation. In particular, on the oppposite to the weak convergence rate, the joint strong convergence rate of the nonrecursive estimators θ_n^* and $\tilde{\mu}_n^*$ cannot be obtained by following the lines of the proof of Theorem 2, and remains an open question.

3 Proofs

Let us first note that an important consequence of (A3) which will be used throughout the proofs is that

$$\text{if } \beta a < 1, \text{ then } \lim_{n\to\infty} \frac{1}{nh_n^\beta} \sum_{i=1}^n h_i^\beta = \frac{1}{1-a\beta}. \tag{9}$$

Moreover, for all $\varepsilon > 0$ small enough,

$$\frac{1}{n}\sum_{i=1}^n h_i^q = O\left(h_n^{q-\varepsilon} + \frac{1}{n}\right). \tag{10}$$

As a matter of fact: (i) if $aq < 1$, (10) follows easily from (9); (ii) if $aq > 1$, since $\sum_i h_i^q$ is summable, (10) holds; (iii) if $aq = 1$, since $a(q-\varepsilon) < 1$, using (9) again, we have $n^{-1}\sum_{i=1}^n h_i^q = O(h_n^{q-\varepsilon})$, and thus (10) follows. Of course (9) and (10) also hold when (h_n) and a are replaced by $(\tilde{h}_n)$ and $\tilde{a}$, respectively.

Our proofs are now organized as follows. Section 3.1 is devoted to the proof of the strong consistency of θ_n and $\tilde{\mu}_n$. In Section 3.2, we give the convergence rate of the derivatives of f_n. In Section 3.3, we show how the study of the joint weak and strong convergence rate of θ_n and $\tilde{\mu}_n$ can be related to the one of $\nabla f_n(\theta)$ and $\tilde{f}_n(\theta)$. In Section 3.4 (respectively in Section 3.5), we establish the joint weak convergence rate (respectively the joint strong convergence rate) of $\nabla f_n(\theta)$ and $\tilde{f}_n(\theta)$. Finally, Section 3.6 is devoted to the proof of Theorems 1 and 2.

3.1 Proof of Proposition 1

Since θ_n is the mode of f_n and θ the mode of f, we have:

$$\begin{aligned} 0 \le f(\theta) - f(\theta_n) &= [f(\theta) - f_n(\theta_n)] + [f_n(\theta_n) - f(\theta_n)] \le [f(\theta) - f_n(\theta)] + [f_n(\theta_n) - f(\theta_n)] \\ &\le \big|f(\theta) - f_n(\theta)\big| + \big|f_n(\theta_n) - f(\theta_n)\big| \le 2\|f_n - f\|_\infty. \end{aligned} \tag{11}$$

The application of Theorem 5 in Mokkadem et al. [15] with $|\alpha| = 0$ and $v_n = \log n$ ensures that for any $\delta > 0$, there exists $c(\delta) > 0$ such that $\mathbb{P}[(\log n)\|f_n - \mathbb{E}(f_n)\|_\infty \ge \delta] \le \exp(-c(\delta)\sum_{i=1}^n h_i^d/(\log n)^2)$. In view of (9), since $ad < 1$, we can write

$$n^2 \exp\left(-c(\delta)\frac{\sum_{i=1}^n h_i^d}{(\log n)^2}\right) = n^2 \exp\left(-c(\delta)\frac{nh_n^d}{(\log n)^2}\frac{\sum_{i=1}^n h_i^d}{nh_n^d}\right) = o(1).$$

Borell-Cantelli's Lemma ensures that $\lim_{n\to\infty}\|f_n - \mathbb{E}(f_n)\|_\infty = 0$ a.s. Since $\lim_{n\to\infty}\|\mathbb{E}(f_n) - f\|_\infty = 0$, it follows from (11) that $\lim_{n\to\infty} f(\theta_n) = f(\theta)$ a.s. Since f is continuous, $\lim_{\|z\|\to\infty} f(z) = 0$ and θ is the unique mode of f, we deduce that $\lim_{n\to\infty}\theta_n = \theta$ a.s. Now, we have

$$|\tilde{\mu}_n - \mu| \le |\tilde{f}_n(\theta_n) - f(\theta_n)| + |f(\theta_n) - f(\theta)| \le \|\tilde{f}_n - f\|_\infty + 2\|f_n - f\|_\infty,$$

where the last inequality follows from (11). As previously, one can show that $\lim_{n\to\infty}\|\tilde{f}_n - f\|_\infty = 0$ and thus $\lim_{n\to\infty}\tilde{\mu}_n = \mu$ a.s.

3.2 Convergence rate of the derivatives of the density

For any d-uplet $[\alpha] = \Big(\alpha_1, \ldots, \alpha_d\Big) \in \mathbb{N}^d$, we set $|\alpha| = \alpha_1 + \cdots + \alpha_d$ and, for any function g, let $\partial^{[\alpha]}g(x) = \partial^{|\alpha|}g/(\partial x_1^{\alpha_1}\ldots\partial x_d^{\alpha_d})(x)$ denote the $[\alpha]$-th partial derivative of g.

Lemma 1 *Assume (A3)-(A5) hold. Let (g_n) and (b_n) be defined as follows:*

$$\left\{ \begin{array}{l} g_n = f_n \quad and \quad b_n = h_n \quad or \\ g_n = \tilde{f}_n \quad and \quad b_n = \tilde{h}_n. \end{array} \right. \tag{12}$$

For $|\alpha| \in \{0,1,2\}$, we have

$$\lim_{n\to\infty} \frac{n}{\sum_{i=1}^n b_i^q} \left[\mathbb{E}\left[\partial^{[\alpha]} g_n(x)\right] - \partial^{[\alpha]} f(x) \right] \;=\; \frac{(-1)^q}{q!} \partial^{[\alpha]} \left(\sum_{j=1}^d \beta_j^q \frac{\partial^q f}{\partial x_j^q} \right)(x)$$

where β_j^q is defined in (5). Moreover, if we set $M_q = \sup_{x\in\mathbb{R}^d} \|D^q \partial^{[\alpha]} f(x)\|$, then

$$\lim_{n\to\infty} \frac{n}{\sum_{i=1}^n b_i^q} \sup_{x\in\mathbb{R}^d} \left| \mathbb{E}\left(\partial^{[\alpha]} g_n(x)\right) - \partial^{[\alpha]} f(x) \right| \;\leq\; \frac{M_q}{q!} \int_{\mathbb{R}^d} \|z\|^q \, |K(z)| \, dz.$$

Lemma 2 *Let U be a compact set of $\mathbb{R}^d$ and assume that (A1)iii), (A3), (A4) and (A5)ii) hold. Let (g_n) and (b_n) be defined as in (12). Then, for all $\gamma > 0$ and $|\alpha| = 1, 2$, we have*

$$\sup_{x\in U} \left| \partial^{[\alpha]} g_n(x) - \mathbb{E}\left(\partial^{[\alpha]} g_n(x)\right) \right| \;=\; O\left(\sqrt{\frac{(\log n)^{1+\gamma}}{\sum_{i=1}^n b_i^{d+2|\alpha|}}} \right) \quad a.s.$$

Lemma 1 is proved in Mokkadem et al. [15]. We now prove Lemma 2. Set $v_n = [\sum_{i=1}^n b_i^{d+2|\alpha|}]^{1/2} [(\log n)^{1+\gamma}]^{-1/2}$. Applying Proposition 3 in Mokkadem et al. [15], it holds that for any $\delta > 0$, there exists $c(\delta) > 0$ such that

$$\mathbb{P}\left[\sup_{x\in U} v_n \left| \partial^{[\alpha]} g_n(x) - \mathbb{E}\left(\partial^{[\alpha]} g_n(x)\right) \right| \geq \delta \right] \;\leq\; \exp\left(-c(\delta) \frac{\sum_{i=1}^n b_i^{d+2|\alpha|}}{2 v_n^2} \right).$$

Since $\lim_{n\to\infty} \sum_{i=1}^n b_i^{d+2|\alpha|} / (v_n^2 \log n) = \infty$ we have, for n large enough, $c(\delta) \sum_{i=1}^n b_i^{d+2|\alpha|} / (2v_n^2) \geq 2 \log n$, and Lemma 2 follows from the application of Borel-Cantelli's Lemma.

3.3 Relationship between $((\boldsymbol{\theta_n - \theta})^T, (\tilde{\boldsymbol{\mu}}_n - \boldsymbol{\mu}))^T$ and $([\nabla \boldsymbol{f_n}(\boldsymbol{\theta})]^T, \tilde{\boldsymbol{f}}_n(\boldsymbol{\theta}) - \boldsymbol{f}(\boldsymbol{\theta}))^T$

By definition of θ_n, we have $\nabla f_n(\theta_n) = 0$ so that

$$\nabla f_n(\theta_n) - \nabla f_n(\theta) = -\nabla f_n(\theta). \tag{13}$$

For each $i \in \{1, \ldots, d\}$, a Taylor expansion applied to the real valued application $\partial f_n / \partial x_i$ implies the existence of $\varepsilon_n(i) = (\varepsilon_n^{(1)}(i), \ldots, \varepsilon_n^{(d)}(i))^t$ such that

$$\left\{ \begin{array}{l} \frac{\partial f_n}{\partial x_i}(\theta_n) - \frac{\partial f_n}{\partial x_i}(\theta) = \sum_{j=1}^d \frac{\partial^2 f_n}{\partial x_i \partial x_j}(\varepsilon_n(i)) \left(\theta_n^{(j)} - \theta^{(j)} \right), \\ \left| \varepsilon_n^{(j)}(i) - \theta^{(j)} \right| \leq \left| \theta_n^{(j)}(i) - \theta^{(j)} \right| \quad \forall j \in \{1, \ldots, d\}. \end{array} \right.$$

Define the $d \times d$ matrix $H_n = (H_n^{(i,j)})_{1\leq i,j\leq d}$ by setting $H_n^{(i,j)} = \frac{\partial^2 f_n}{\partial x_i \partial x_j}(\varepsilon_n(i))$; Equation (13) can be then rewritten as $H_n(\theta_n - \theta) = -\nabla f_n(\theta)$. Now, set

$$R_n \;=\; \tilde{f}_n(\theta_n) - \tilde{f}_n(\theta). \tag{14}$$

We can then write:

$$\left(\begin{array}{c} \left[D^2 f(\theta)\right]^{-1} H_n(\theta_n - \theta) \\ \tilde{\mu}_n - \mu \end{array} \right) = \left(\begin{array}{c} -\left[D^2 f(\theta)\right]^{-1} \nabla f_n(\theta) \\ \tilde{f}_n(\theta) - f(\theta) \end{array} \right) + \left(\begin{array}{c} 0 \\ R_n \end{array} \right). \quad (15)$$

Let U be a compact set of $\mathbb{R}^d$ containing θ. The combination of Lemmas 1 and 2 with $|\alpha| = 2$, $g_n = f_n$ and $b_n = h_n$ ensures that for any $\gamma > 0$ and $\varepsilon > 0$ small enough,

$$\begin{aligned} \sup_{x \in U} \left| \partial^{[\alpha]} f_n(x) - \partial^{[\alpha]} f(x) \right| &= O\left(\sqrt{\frac{(\log n)^{1+\gamma}}{\sum_{i=1}^n h_i^{d+4}}} + \frac{\sum_{i=1}^n h_i^q}{n} \right) \text{ a.s.} \\ &= O\left(\sqrt{\frac{(\log n)^{1+\gamma}}{n h_n^{d+4}}} + h_n^{q-\varepsilon} + \frac{1}{n} \right) = o(1) \text{ a.s.} \quad (16) \end{aligned}$$

Since $D^2 f$ is continuous in a neighbourhood of θ and since $\lim_{n\to\infty} \theta_n = \theta$ a.s., (16) ensures that $\lim_{n\to\infty} H_n = D^2 f(\theta)$ a.s. It follows that the weak and a.s. behaviours of $((\theta_n - \theta)^T, (\tilde{\mu}_n - \mu))^T$ are given by the one of the right-hand-sided term of (15).

3.4 Weak convergence rate of $\left([\nabla f_n(\theta)]^T, \tilde{f}_n(\theta) - f(\theta)\right)^T$

Let us at first assume that the following lemma holds.

Lemma 3 *Let Assumptions (A1)i), (A1)iv), (A3), (A4)i) and (A4)ii) hold. Then*

$$W_n = \left(\begin{array}{c} \sqrt{n h_n^{d+2}} \Big[\nabla f_n(\theta) - \mathbb{E}\big(\nabla f_n(\theta)\big) \Big] \\ \sqrt{n \tilde{h}_n^d} \Big[\tilde{f}_n(\theta) - \mathbb{E}\big(\tilde{f}_n(\theta)\big) \Big] \end{array} \right) \xrightarrow{\mathcal{D}} \mathcal{N}\Big(0, \Sigma\Big).$$

The application of Lemma 1 gives

$$\lim_{n\to\infty} \left(\begin{array}{c} \frac{n}{\sum_{i=1}^n h_i^q} \mathbb{E}\big(\nabla f_n(\theta)\big) \\ \frac{n}{\sum_{i=1}^n \tilde{h}_i^q} \Big[\mathbb{E}\big(\tilde{f}_n(\theta)\big) - f(\theta) \Big] \end{array} \right) = \left(\begin{array}{c} \frac{(-1)^q}{q!} \nabla \Big(\sum_{j=1}^d \beta_j^q \frac{\partial^q f}{\partial x_j^q}(\theta) \Big) \\ \frac{(-1)^q}{q!} \sum_{j=1}^d \beta_j^q \frac{\partial^q f}{\partial x_j^q}(\theta) \end{array} \right). \quad (17)$$

1) If $aq < 1$ and $\tilde{a}q < 1$, by using (9), it is straightforward to see that

$$\lim_{n\to\infty} \left(\begin{array}{c} \frac{1}{h_n^q} \mathbb{E}\big(\nabla f_n(\theta)\big) \\ \frac{1}{\tilde{h}_n^q} \Big[\mathbb{E}\big(\tilde{f}_n(\theta)\big) - f(\theta) \Big] \end{array} \right) = B_q(\theta). \quad (18)$$

2) Let us now consider the case $aq \geq 1$ and $\tilde{a}q \geq 1$. We have

$$\sqrt{n h_n^{d+2}} \mathbb{E}\left(\nabla f_n(\theta)\right) = \sqrt{n h_n^{d+2}} \frac{\sum_{i=1}^n h_i^q}{n} \frac{n}{\sum_{i=1}^n h_i^q} \mathbb{E}\left(\nabla f_n(\theta)\right),$$

with, in view of (10), for all $\varepsilon > 0$ small enough,

$$\sqrt{n h_n^{d+2}} \frac{\sum_{i=1}^n h_i^q}{n} = O\left(n^{\frac{1}{2}(1-(a-\varepsilon)(d+2))} n^{-aq+a\varepsilon} \right) = o(1).$$

Applying (17), it follows that $\lim_{n\to\infty} \sqrt{n h_n^{d+2}} \mathbb{E}(\nabla f_n(\theta)) = 0$. Proceeding in the same way for $\mathbb{E}(\tilde{f}_n(\theta))$, we obtain

$$\lim_{n\to\infty} \left(\begin{array}{c} \sqrt{n h_n^{d+2}} \mathbb{E}\big(\nabla f_n(\theta)\big) \\ \sqrt{n \tilde{h}_n^d} \Big[\mathbb{E}\big(\tilde{f}_n(\theta)\big) - f(\theta) \Big] \end{array} \right) = 0. \quad (19)$$

The combination of either (18) or (19) and of Lemma 3 gives the weak convergence rate of $([\nabla f_n(\theta)]^T, \tilde{f}_n(\theta) - f(\theta))^T$:

- If (C1) holds, then

$$\begin{pmatrix} \sqrt{nh_n^{d+2}}\nabla f_n(\theta) \\ \sqrt{n\tilde{h}_n^d}(\tilde{f}_n(\theta) - f(\theta)) \end{pmatrix} \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Sigma). \tag{20}$$

- If $a = (d+2q+2)^{-1}$, $\tilde{a} = (d+2q)^{-1}$, and if there exist $c, \tilde{c} \geq 0$ such that $\lim_{n\to\infty} nh_n^{d+2q+2} = c$ and $\lim_{n\to\infty} n\tilde{h}_n^{d+2q} = \tilde{c}$, then

$$\begin{pmatrix} \sqrt{nh_n^{d+2}}\nabla f_n(\theta) \\ \sqrt{n\tilde{h}_n^d}(\tilde{f}_n(\theta) - f(\theta)) \end{pmatrix} \xrightarrow{\mathcal{D}} \mathcal{N}(D(c, \tilde{c})B_q(\theta), \Sigma). \tag{21}$$

- If (C2) holds, since $aq < 1$ and $\tilde{a}q < 1$, (9) implies that

$$\begin{pmatrix} \frac{1}{h_n^q}\nabla f_n(\theta) \\ \frac{1}{\tilde{h}_n^q}(\tilde{f}_n(\theta) - f(\theta)) \end{pmatrix} \xrightarrow{\mathbb{P}} B_q(\theta). \tag{22}$$

Proof of Lemma 3 To prove Lemma 3, we first prove that

$$\lim_{n\to\infty} \mathbb{E}\Big(W_n W_n^T\Big) = \Sigma, \tag{23}$$

and then check that (W_n) satisfies Lyapounov's condition. Set

$$\begin{aligned} Y_{k,n} &= \frac{1}{\sqrt{nh_n^{-d-2}}} h_k^{-d-1}\left[\nabla K\left(\frac{\theta - X_k}{h_k}\right) - \mathbb{E}\left(\nabla K\left(\frac{\theta - X_k}{h_k}\right)\right)\right] \\ Z_{k,n} &= \frac{1}{\sqrt{n\tilde{h}_n^{-d}}} \tilde{h}_k^{-d}\left[K\left(\frac{\theta - X_k}{\tilde{h}_k}\right) - \mathbb{E}\left(K\left(\frac{\theta - X_k}{\tilde{h}_k}\right)\right)\right], \end{aligned}$$

and note that

$$\mathbb{E}\Big(W_n W_n^T\Big) = \sum_{k=1}^n \begin{pmatrix} \mathbb{E}\Big(Y_{k,n}Y_{k,n}^T\Big) & \mathbb{E}\Big(Y_{k,n}Z_{k,n}\Big) \\ \mathbb{E}\Big(Y_{k,n}^T Z_{k,n}\Big) & \mathbb{E}\Big(Z_{k,n}^2\Big) \end{pmatrix}.$$

Now, for any $s, t \in \{1, \ldots, d\}$, we have

$$\begin{aligned} \mathbb{E}\left[\frac{\partial K}{\partial x_s}\left(\frac{\theta - X_k}{h_k}\right)\frac{\partial K}{\partial x_t}\left(\frac{\theta - X_k}{h_k}\right)\right] &= \int_{\mathbb{R}^d} \frac{\partial K}{\partial x_s}\left(\frac{\theta - y}{h_k}\right)\frac{\partial K}{\partial x_t}\left(\frac{\theta - y}{h_k}\right) f(y)dy \\ &= h_k^d f(\theta)G_{s,t} + o(h_k^d), \end{aligned}$$

and since, $\mathbb{E}\left[\frac{\partial K}{\partial x_s}\left(\frac{\theta - X_k}{h_k}\right)\right] = O(h_k^d)$, we deduce that

$$\begin{aligned} &\mathbb{E}\left(\left[\nabla K\left(\tfrac{\theta - X_k}{h_k}\right) - \mathbb{E}\left(\nabla K\left(\tfrac{\theta - X_k}{h_k}\right)\right)\right]\left[\nabla K\left(\tfrac{\theta - X_k}{h_k}\right) - \mathbb{E}\left(\nabla K\left(\tfrac{\theta - X_k}{h_k}\right)\right)\right]^T\right) \\ &\qquad = f(\theta)Gh_k^d\Big[1 + o(1)\Big] \end{aligned} \tag{24}$$

which implies that $\lim_{n\to\infty}\sum_{k=1}^n \mathbb{E}(Y_{k,n}Y_{k,n}^T) = f(\theta)[1+a(d+2)]^{-1}G$. In the same way, we have

$$\mathbb{E}\left(\left[K\left(\frac{\theta-X_k}{\tilde{h}_k}\right)-\mathbb{E}\left(K\left(\frac{\theta-X_k}{\tilde{h}_k}\right)\right)\right]^2\right) \;=\; \tilde{h}_k^d f(\theta)\int_{\mathbb{R}^d}K^2(z)dz\Big[1+o(1)\Big] \tag{25}$$

and thus $\lim_{n\to\infty}\sum_{k=1}^n \mathbb{E}(Z_{k,n}^2) = f(\theta)[1+\tilde{a}d]^{-1}\int_{\mathbb{R}^d}K^2(z)dz$. Moreover, set $h_n^* = \min(h_n,\tilde{h}_n)$; we have

$$\mathbb{E}\left[\nabla K\left(\frac{\theta-X_k}{h_k}\right)K\left(\frac{\theta-X_k}{\tilde{h}_k}\right)\right] \;=\; h_k^{*d}\int_{\mathbb{R}^d}\nabla K\Big(\frac{h_k^*}{h_k}z\Big)K\Big(\frac{h_k^*}{\tilde{h}_k}z\Big)f(\theta-h_k^*z)dz.$$

Noting that $f(\theta-h_k^*z) = f(\theta)+h_k^*R_k(\theta,z)$ with $|R_k(\theta,z)|\le\|\nabla f\|_\infty\|z\|$, we get

$$\begin{aligned}&\mathbb{E}\left[\nabla K\left(\frac{\theta-X_k}{h_k}\right)K\left(\frac{\theta-X_k}{\tilde{h}_k}\right)\right]\\ &= h_k^{*d}\left[f(\theta)\int_{\mathbb{R}^d}\nabla K\Big(\frac{h_k^*}{h_k}z\Big)K\Big(\frac{h_k^*}{\tilde{h}_k}z\Big)dz + h_k^*\int_{\mathbb{R}^d}\nabla K\Big(\frac{h_k^*}{h_k}z\Big)K\Big(\frac{h_k^*}{\tilde{h}_k}z\Big)R_k(\theta,z)dz\right].\end{aligned}$$

Since the function $z\mapsto\big[\nabla K(z)\big]K($z is odd (in each coordinate), the first right-handed integral is zero, and, since h_k^* equals either h_k or $\tilde{h}_k$, we get

$$\begin{aligned}&\left\|\mathbb{E}\left[\nabla K\left(\frac{\theta-X_k}{h_k}\right)K\left(\frac{\theta-X_k}{\tilde{h}_k}\right)\right]\right\|\\ &\le\; h_k^{*(d+1)}\|\nabla f\|_\infty\left[\|K\|_\infty\int_{\mathbb{R}^d}\|z\|\|\nabla K(z)\|dz+\|\nabla K\|_\infty\int_{\mathbb{R}^d}\|z\||K(z)|dz\right] = O\left(h_k^{*(d+1)}\right).\end{aligned}$$

We then deduce that

$$\begin{aligned}&\mathbb{E}\left(\left[\nabla K\left(\frac{\theta-X_k}{h_k}\right)-\mathbb{E}\left(\nabla K\left(\frac{\theta-X_k}{h_k}\right)\right)\right]\left[K\left(\frac{\theta-X_k}{\tilde{h}_k}\right)-\mathbb{E}\left(K\left(\frac{\theta-X_k}{\tilde{h}_k}\right)\right)\right]\right)\\ =\;& O\left(\Big[\min(h_k,\tilde{h}_k)\Big]^{d+1}\right)+O\left(h_k^d\tilde{h}_k^d\right) = O\left(h_k^{\frac{d+1}{2}}\tilde{h}_k^{\frac{d+1}{2}}\right),\end{aligned} \tag{26}$$

and thus, in view of (9),

$$\sum_{k=1}^n\mathbb{E}\Big(Y_{k,n}Z_{k,n}\Big) = O\left(\frac{1}{\sqrt{(nh_n^{-d-2})(n\tilde{h}_n^{-d})}}\sum_{k=1}^n h_k^{-\frac{d+1}{2}}\tilde{h}_k^{\frac{1-d}{2}}\right) = o(1),$$

which concludes the proof of (23). Now we check that (W_n) satisfies the Lyapounov's condition. Set $p>2$. Since K and ∇K are bounded and integrable, we have $\int_{\mathbb{R}^d}\|\nabla K(z)\|^p dz<\infty$ and $\int_{\mathbb{R}^d}|K(z)|^p dz<\infty$. It follows that

$$\begin{aligned}\sum_{k=1}^n\mathbb{E}\left(\|Y_{k,n}\|^p\right) &= O\left(\frac{1}{(nh_n^{-d-2})^{\frac{p}{2}}}\sum_{k=1}^n h_k^{(-d-1)p}\int_{\mathbb{R}^d}\left\|\nabla K\left(\frac{\theta-y}{h_k}\right)\right\|^p f(y)dy\right)\\ &= O\left(\frac{1}{(nh_n^{-d-2})^{\frac{p}{2}}}\sum_{k=1}^n h_k^{(-d-1)p}h_k^d\right) = o(1),\\ \sum_{k=1}^n\mathbb{E}\left(\big|Z_{k,n}\big|^p\right) &= O\left(\frac{1}{(n\tilde{h}_n^{-d})^{\frac{p}{2}}}\sum_{k=1}^n \tilde{h}_k^{-dp}\int_{\mathbb{R}^d}\left|K\left(\frac{\theta-y}{\tilde{h}_k}\right)\right|^p f(y)dy\right)\\ &= O\left(\frac{1}{(n\tilde{h}_n^{-d})^{\frac{p}{2}}}\sum_{k=1}^n \tilde{h}_k^{-dp}\tilde{h}_k^d\right) = o(1),\end{aligned}$$

which concludes the proof of Lemma 3.

3.5 A.s. convergence rate of $\left([\nabla f_n(\theta)]^T, \tilde{f}_n(\theta) - f(\theta)\right)^T$

Let us at first assume that the following lemma holds.

Lemma 4 *Let Assumptions (A1)i), (A1)iv), (A3), (A4)i), (A4)ii) and (A6) hold. With probability one, the sequence*

$$\frac{1}{\sqrt{2\log\log n}} \begin{pmatrix} \sqrt{nh_n^{d+2}}\big[\nabla f_n(\theta) - \mathbb{E}\big(\nabla f_n(\theta)\big)\big] \\ \sqrt{n\tilde{h}_n^d}\left[\tilde{f}_n(\theta) - \mathbb{E}\big(\tilde{f}_n(\theta)\big)\right] \end{pmatrix}$$

is relatively compact and its limit set is $\mathcal{E} = \big\{\nu \in \mathbb{R}^{d+1} \text{ such that } \nu^T\Sigma^{-1}\nu \leq 1\big\}$.

The combination of either (18) or (19) and of Lemma 4 gives the almost sure convergence rate of $\left([\nabla f_n(\theta)]^T, \tilde{f}_n(\theta) - f(\theta)\right)^T$:

- If (C1) holds, then, with probability one, the sequence

$$\frac{1}{\sqrt{2\log\log n}} \begin{pmatrix} \sqrt{nh_n^{d+2}}\nabla f_n(\theta) \\ \sqrt{n\tilde{h}_n^d}\left[\tilde{f}_n(\theta) - f(\theta)\right] \end{pmatrix} \tag{27}$$

  is relatively compact and its limit set is $\mathcal{E} = \big\{\nu \in \mathbb{R}^{d+1} \text{ such that } \nu^T\Sigma^{-1}\nu \leq 1\big\}$.

- If $a = (d+2q+2)^{-1}$, $\tilde{a} = (d+2q)^{-1}$, and if there exist $c, \tilde{c} \geq 0$ such that $\lim_{n\to\infty} nh_n^{d+2q+2}/(2\log\log n) = c$ and $\lim_{n\to\infty} n\tilde{h}_n^{d+2q}/(2\log\log n) = \tilde{c}$, then with probability one, the sequence

$$\frac{1}{\sqrt{2\log\log n}} \begin{pmatrix} \sqrt{nh_n^{d+2}}\nabla f_n(\theta) \\ \sqrt{n\tilde{h}_n^d}\left[\tilde{f}_n(\theta) - f(\theta)\right] \end{pmatrix} \tag{28}$$

  is relatively compact and its limit set is

$$\mathcal{E} = \Big\{\nu \in \mathbb{R}^{d+1} \text{ such that } \left(\nu - D(c,\tilde{c})B_q(\theta)\right)^T \Sigma^{-1} \left(\nu - D(c,\tilde{c})B_q(\theta)\right) \leq 1\Big\}.$$

- If (C'2) holds, then

$$\begin{pmatrix} \frac{1}{h_n^q}\nabla f_n(\theta) \\ \frac{1}{\tilde{h}_n^q}[\tilde{f}_n(\theta) - f(\theta)] \end{pmatrix} \xrightarrow{\text{a.s.}} B_q(\theta). \tag{29}$$

We now prove Lemma 4. Set

$$\Gamma = f(\theta) \begin{pmatrix} G & 0 \\ 0 & \int_{\mathbb{R}^d} K^2(z)dz \end{pmatrix}, \ \Delta_n = \begin{pmatrix} \frac{1}{\sqrt{nh_n^{-d-2}}} I_d & 0 \\ 0 & \frac{1}{\sqrt{n\tilde{h}_n^{-d}}} \end{pmatrix}, \ Q_n = \begin{pmatrix} \sqrt{h_n^{-d-2}} I_d & 0 \\ 0 & \sqrt{\tilde{h}_n^{-d}} \end{pmatrix},$$

let (ε_n) be a sequence of $\mathbb{R}^{d+1}$-valued, independent and $\mathcal{N}(0,\Gamma)$-distributed random vectors, and set $S_n = \sum_{k=1}^n Q_k\varepsilon_k$. In order to prove Lemma 4, we first establish the following Lemma 5 in Section 3.5.1, and then show in Section 3.5.2 how Lemma 4 can be deduced from Lemma 5.

Lemma 5 *Let Assumptions (A1)i), (A1)iv), (A3), (A4)i), (A4)ii) and (A6)ii) hold. With probability one, the sequence* $(T_n) \equiv (\Sigma^{-1/2}\Delta_n S_n/\sqrt{2\log\log n})$ *is relatively compact and its limit set is the unit ball* $\overline{\mathcal{B}}_{d+1}(0,1) = \Big\{\nu \in \mathbb{R}^{d+1} \text{ such that } \|\nu\|_2 \leq 1\Big\}$.

3.5.1 Proof of Lemma 5

Set $B_n = \mathbb{E}(S_n S_n^T)$, let $\|x\|_2$ (respectively $|\!|\!|A|\!|\!|_2$) denote the euclidean norm (respectively the spectral norm) of the vector x (respectively of the matrix A). The application of Theorem 2 in Koval [11] ensures that

$$\limsup_{n\to\infty} \frac{\|\Sigma^{-1/2}\Delta_n S_n\|_2}{\sqrt{2\,|\!|\!|\Sigma^{-1/2}\Delta_n B_n \Delta_n \Sigma^{-1/2}|\!|\!|_2 \log\log |\!|\!| B_n |\!|\!|_2}} \leq 1 \quad \text{a.s.}$$

Since $\lim_{n\to\infty} \Delta_n B_n \Delta_n = \Sigma$ and $\log\log |\!|\!| B_n |\!|\!|_2 \sim \log\log n$, we deduce that

$$\limsup_{n\to\infty} \|T_n\|_2 \leq 1 \quad \text{a.s.} \tag{30}$$

Thus, the sequence (T_n) is relatively compact and its limit set $\mathcal{U}$ is included in $\overline{\mathcal{B}}_{d+1}(0,1)$. Now, set $\mathcal{S}_{d+1} = \Big\{ w \in \mathbb{R}^{d+1},\ \|w\|_2 = 1 \Big\}$, and let us at first assume that

$$\forall w \in \mathcal{S}_{d+1},\quad \limsup_{n\to\infty} w^T T_n \geq 1 \quad \text{a.s.} \tag{31}$$

The combination of (30) and (31) ensures that, with probability one, $\forall \varepsilon > 0, \forall n_0 \geq 1, \exists n \geq n_0$ such that $w^T T_n > 1 - \varepsilon$ and $\|T_n\|_2^2 \leq 1 + \varepsilon$. Noting that $\|T_n - w\|_2^2 = \|T_n\|_2^2 + \|w\|_2^2 - 2w^T T_n$, it follows that, with probability one, $\forall \varepsilon > 0$, $\forall n_0 \geq 1$, $\exists n \geq n_0$ such that $\|T_n - w\|_2^2 \leq 1+\varepsilon+1-2(1-\varepsilon) = 3\varepsilon$. Thus, with probability one, $\mathcal{S}_{d+1} \subset \mathcal{U}$. To deduce that $\overline{\mathcal{B}}_{d+1}(0,1) \subset \mathcal{U}$, we introduce (e_k), a sequence of real-valued, independent, and $\mathcal{N}(0,1)$-distributed random variables such that (e_k) is independent of (ε_k). Moreover, we set

$$\tilde{Q}_n = \begin{pmatrix} \sqrt{h_n^{-d-2}} I_{d+1} & 0 \\ 0 & \sqrt{\tilde{h}_n^{-d}} \end{pmatrix}, \quad \tilde{S}_n = \sum_{k=1}^{n} \tilde{Q}_k \begin{pmatrix} e_k \\ \varepsilon_k \end{pmatrix}$$

$$\tilde{\Delta}_n = \begin{pmatrix} \frac{1}{\sqrt{n h_n^{-d-2}}} I_{d+1} & 0 \\ 0 & \frac{1}{\sqrt{n \tilde{h}_n^{-d}}} \end{pmatrix}, \quad \text{and} \quad \tilde{\Sigma} = \begin{pmatrix} 1 & 0 \\ 0 & \Sigma \end{pmatrix}.$$

We then note that the previous result applied to $(\tilde{T}_n) \equiv (\tilde{\Sigma}^{-1/2}\tilde{\Delta}_n \tilde{S}_n / \sqrt{2\log\log n})$ ensures that, with probability one, $\mathcal{S}_{d+2} = \{w \in \mathbb{R}^{d+2},\ \|w\|_2 = 1\}$ is included in the limit set of $\tilde{T}_n$. Now let $\pi : \mathbb{R}^{d+2} \longrightarrow \mathbb{R}^{d+1}$ be the projection map defined by $\pi((x_1, \ldots, x_{d+2})^T) = (x_2, \ldots, x_{d+2})^T$. We clearly have $\pi(\mathcal{S}_{d+2}) = \overline{\mathcal{B}}_{d+1}(0,1)$ and $\pi(\tilde{T}_n) = T_n$, and thus deduce that, with probability one, $\overline{\mathcal{B}}_{d+1}(0,1)$ is included in the limit set of T_n. To conclude the proof of Lemma 5, it remains to prove (31). In fact, we shall prove that,

$$\forall w \neq 0,\quad \limsup_{n\to\infty} \frac{w^T \Delta_n S_n}{\sqrt{2\log\log n}} \geq \sqrt{w^T \Sigma w} \quad \text{a.s.} \tag{32}$$

Set $v_n = \min\{[n h_n^{-(d+2)}]^{1/2}; [n\tilde{h}_n^{-d}]^{1/2}\}$, $A_n = v_n w^T \Delta_n$ and $V_n = \mathbb{E}\big(A_n S_n S_n^T A_n^T\big)$; we follow a method used by Petrov [19] in the proof of his Theorems 7.1 and 7.2. Since $\lim_{n\to\infty} V_n = \infty$, $\forall \tau > 0$, there exists a non-decreasing sequence of integers n_k such that $n_k \to \infty$ as $k \to \infty$ and $V_{n_k - 1} \leq (1+\tau)^k \leq V_{n_k}$, $(k = 1, 2, \ldots)$. Since $\lim_{n\to\infty} V_{n-1}/V_n = 1$, we obtain $V_{n_k} \sim (1+\tau)^k$. Moreover, we have

$$V_{n_k} - V_{n_{k-1}} = V_{n_k}\big(1 - \frac{V_{n_{k-1}}}{V_{n_k}}\big) \sim V_{n_k} \frac{\tau}{\tau + 1}. \tag{33}$$

Set

$$\chi(n) = \sqrt{2V_n \log\log V_n}, \quad \psi(n_k) = \sqrt{2(V_{n_k} - V_{n_{k-1}}) \log\log(V_{n_k} - V_{n_{k-1}})}.$$

It follows from (33) that $\psi(n_k) \sim \tau^{1/2}\chi(n_{k-1})$. Then for any $\gamma \in]0,1[$ and k sufficiently large, we have

$$\begin{aligned}
&\mathbb{P}\left(A_{n_k}S_{n_k} - A_{n_k}S_{n_{k-1}} \geq (1-\gamma)\psi(n_k)\right) \\
&\quad \geq \mathbb{P}\left(A_{n_k}S_{n_k} \geq (1-\frac{\gamma}{2})\psi(n_k)\right) - \mathbb{P}\left(A_{n_k}S_{n_{k-1}} \geq \frac{\gamma\psi(n_k)}{2}\right) \\
&\quad \geq \mathbb{P}\left(A_{n_k}S_{n_k} \geq (1-\frac{\gamma}{2})\chi(n_k)\right) - \mathbb{P}\left(A_{n_k}S_{n_{k-1}} \geq \frac{\gamma\sqrt{\tau}}{3}\chi(n_{k-1})\right). \qquad (34)
\end{aligned}$$

Since $A_{n_k}S_{n_k}$ is $\mathcal{N}\big(0, V_{n_k}\big)$-distributed, we have

$$\begin{aligned}
\mathbb{P}\left(A_{n_k}S_{n_k} \geq (1-\frac{\gamma}{2})\chi(n_k)\right) &= \frac{1}{\sqrt{2\pi}} \int_{(1-\frac{\gamma}{2})\sqrt{2V_{n_k}\log\log V_{n_k}}}^{\infty} \exp\left(-\frac{t^2}{2}\right) dt \\
&\geq \left[\log V_{n_k}\right]^{-(1+\mu)(1-\frac{\gamma}{2})^2} \qquad (35)
\end{aligned}$$

for every μ and sufficiently large k. Set $\tilde{V}_{n_k} = v_{n_k}^2 w^T \Delta_{n_k} B_{n_{k-1}} \Delta_{n_k} w$; since $A_{n_k}S_{n_{k-1}}$ is $\mathcal{N}\big(0, \tilde{V}_{n_k}\big)$-distributed, we have

$$\mathbb{P}\Big(A_{n_k}S_{n_{k-1}} \geq \frac{\gamma\sqrt{\tau}}{3}\chi(n_{k-1})\Big) = \frac{1}{\sqrt{2\pi}} \int_{\frac{\gamma\sqrt{\tau}}{3}\sqrt{2\frac{V_{n_{k-1}}}{\tilde{V}_{n_k}}\log\log V_{n_{k-1}}}}^{\infty} \exp\left(-\frac{t^2}{2}\right) dt.$$

Let $\rho_{min}(A)$ (respectively $\rho_{max}(A)$) denote the smallest (respectively the largest) eigenvalue of a matrix A, set $\Sigma_n = \Delta_n B_n \Delta_n$, and note that

$$\frac{V_{n_{k-1}}}{\tilde{V}_{n_k}} \geq \frac{v_{n_{k-1}}^2 \rho_{min}(\Sigma_{n_{k-1}})}{v_{n_k}^2 \rho_{max}(\Delta_{n_k}\Delta_{n_{k-1}}^{-1}\Sigma_{n_{k-1}}\Delta_{n_{k-1}}^{-1}\Delta_{n_k})} \qquad (36)$$

with

$$\begin{aligned}
\rho_{max}(\Delta_{n_k}\Delta_{n_{k-1}}^{-1}\Sigma_{n_{k-1}}\Delta_{n_{k-1}}^{-1}\Delta_{n_k}) &\leq \left|\!\left|\!\left|\Sigma_{n_{k-1}}\Delta_{n_{k-1}}^{-1}\Delta_{n_k}\Delta_{n_k}\Delta_{n_{k-1}}^{-1}\right|\!\right|\!\right|_2 \\
&\leq \left|\!\left|\!\left|\Sigma_{n_{k-1}}\right|\!\right|\!\right|_2 \left|\!\left|\!\left|\Delta_{n_{k-1}}^{-1}\Delta_{n_k}\Delta_{n_k}\Delta_{n_{k-1}}^{-1}\right|\!\right|\!\right|_2. \qquad (37)
\end{aligned}$$

It follows from (9) and Assumption A6)ii) that

$$\left|\!\left|\!\left|\Delta_{n_{k-1}}^{-1}\Delta_{n_k}\Delta_{n_k}\Delta_{n_{k-1}}^{-1}\right|\!\right|\!\right|_2 = \max\left\{\frac{n_{k-1}h_{n_{k-1}}^{-(d+2)}}{n_k h_{n_k}^{-(d+2)}}; \frac{n_{k-1}\tilde{h}_{n_{k-1}}^{-d}}{n_k \tilde{h}_{n_k}^{-d}}\right\} \sim \frac{v_{n_{k-1}}^2}{v_{n_k}^2}. \qquad (38)$$

¿From (36), (37) and (38), we deduce that, for sufficiently large k,

$$\frac{V_{n_{k-1}}}{\tilde{V}_{n_k}} \geq \frac{\rho_{min}(\Sigma_{n_{k-1}})}{2\rho_{max}(\Sigma_{n_{k-1}})} \geq \frac{\rho_{min}(\Sigma)}{4\rho_{max}(\Sigma)}$$

and therefore, for sufficiently large k,

$$\begin{aligned}
\mathbb{P}\left(A_{n_k}S_{n_{k-1}} \geq \frac{\gamma\sqrt{\tau}}{3}\chi(n_{k-1})\right) &\leq \frac{1}{\sqrt{2\pi}} \int_{\frac{\gamma\sqrt{\tau}}{6}\sqrt{\frac{\rho_{min}(\Sigma)}{\rho_{max}(\Sigma)}}\sqrt{2V_{n_{k-1}}\log\log V_{n_{k-1}}}}^{\infty} \exp\left(-\frac{t^2}{2}\right) dt \\
&\leq \left[\log V_{n_{k-1}}\right]^{\frac{-\gamma^2\tau\rho_{min}(\Sigma)}{36\rho_{max}(\Sigma)}}. \qquad (39)
\end{aligned}$$

The inequalities (34), (35) and (39) imply that

$$\mathbb{P}\left(A_{n_k}S_{n_k} - A_{n_k}S_{n_{k-1}} \geq (1-\gamma)\psi(n_k)\right) \geq [\log V_{n_k}]^{-(1+\mu)(1-\frac{\gamma}{2})^2} - \left[\log V_{n_{k-1}}\right]^{\frac{-\gamma^2\tau\rho_{min}(\Sigma)}{36\rho_{max}(\Sigma)}}.$$

Thus, for sufficiently large k and τ, there exists $c>0$ such that c does not depend on k and

$$\mathbb{P}\left(A_{n_k}S_{n_k} - A_{n_k}S_{n_{k-1}} \geq (1-\gamma)\psi(n_k)\right) \geq c\Big[k^{-(1+\mu)(1-\frac{\gamma}{2})^2} - k^{-1}\Big].$$

Choosing μ such that $(1+\mu)(1-\gamma/2)^2 < 1$, we get

$$\mathbb{P}\left(A_{n_k}S_{n_k} - A_{n_k}S_{n_{k-1}} \geq (1-\gamma)\psi(n_k)\right) \geq \frac{c}{2}k^{-(1+\mu)(1-\frac{\gamma}{2})^2}$$

and thus $\sum_k \mathbb{P}(A_{n_k}S_{n_k} - A_{n_k}S_{n_{k-1}} \geq (1-\gamma)\psi(n_k)) = \infty$. Applying Borel-Cantelli's Lemma, we obtain

$$\mathbb{P}\left(A_{n_k}S_{n_k} - A_{n_k}S_{n_{k-1}} \geq (1-\gamma)\psi(n_k) \quad \text{i.o.}\right) = 1. \tag{40}$$

Now,

$$\begin{aligned}
\limsup_{k\to\infty}\frac{\left|A_{n_k}S_{n_{k-1}}\right|}{\chi(n_{k-1})} &\leq \limsup_{k\to\infty}\frac{v_{n_k}\|w\|_2\left|\!\left|\!\left|\Delta_{n_k}\Delta_{n_{k-1}}^{-1}\right|\!\right|\!\right|_2\|\Delta_{n_{k-1}}S_{n_{k-1}}\|_2}{\sqrt{2v_{n_{k-1}}^2(w^T\Delta_{n_{k-1}}B_{n_{k-1}}\Delta_{n_{k-1}}w)\log\log V_{n_{k-1}}}} \\
&\leq \limsup_{k\to\infty}\frac{v_{n_k}\|w\|_2\left|\!\left|\!\left|\Delta_{n_k}\Delta_{n_{k-1}}^{-1}\right|\!\right|\!\right|_2\|\Delta_{n_{k-1}}S_{n_{k-1}}\|_2}{\sqrt{2v_{n_{k-1}}^2(w^T\Sigma w)\log\log V_{n_{k-1}}}}.
\end{aligned}$$

Applying Theorem 2 in Koval [11] again, and using the fact that $\lim_{n\to\infty}\Delta_n B_n\Delta_n = \Sigma$, we obtain $\limsup_{n\to\infty}\|\Delta_n S_n\|_2/\sqrt{2\left|\!\left|\!\left|\Sigma\right|\!\right|\!\right|_2\log\log n} \leq 1$ a.s. Therefore,

$$\limsup_{k\to\infty}\frac{\left|A_{n_k}S_{n_{k-1}}\right|}{\chi(n_{k-1})} \leq \limsup_{k\to\infty}\frac{v_{n_k}\|w\|_2\left|\!\left|\!\left|\Delta_{n_k}\Delta_{n_{k-1}}^{-1}\right|\!\right|\!\right|_2\sqrt{\left|\!\left|\!\left|\Sigma\right|\!\right|\!\right|_2}}{v_{n_{k-1}}\sqrt{w^T\Sigma w}} \quad \text{a.s.}$$

Since $\left|\!\left|\!\left|\Delta_{n_k}\Delta_{n_{k-1}}^{-1}\right|\!\right|\!\right|_2 = [\rho_{max}(\Delta_{n_{k-1}}^{-1}\Delta_{n_k}\Delta_{n_k}\Delta_{n_{k-1}}^{-1})]^{1/2} \leq 2v_{n_{k-1}}/v_{n_k}$, for sufficiently large k, we obtain $\limsup_{k\to\infty}\left|A_{n_k}S_{n_{k-1}}\right|/\chi(n_{k-1}) \leq 2\|w\|_2\sqrt{\left|\!\left|\!\left|\Sigma\right|\!\right|\!\right|_2}/\sqrt{w^T\Sigma w}$ a.s. Set $\varepsilon \in]0,1[$ and $\kappa = 2\|w\|_2\sqrt{\left|\!\left|\!\left|\Sigma\right|\!\right|\!\right|_2}/\sqrt{w^T\Sigma w}$. Noting that

$$(1-\gamma)\psi(n_k) - 2\kappa\chi(n_{k-1}) \sim \Big[(1-\gamma)\sqrt{\tau}(1+\tau)^{-1/2} - 2\kappa(1+\tau)^{-1/2}\Big]\chi(n_k),$$

and noting that γ can be chosen sufficiently small and τ sufficiently large so that $(1-\gamma)\sqrt{\tau}(1+\tau)^{-1/2} - 2\kappa(1+\tau)^{-1/2} > 1-\varepsilon$, we obtain

$$\mathbb{P}\left(A_{n_k}S_{n_k} > (1-\varepsilon)\chi(n_k) \quad \text{i.o.}\right) \geq \mathbb{P}\left(A_{n_k}S_{n_k} > (1-\gamma)\psi(n_k) - 2\kappa\chi(n_{k-1}) \quad \text{i.o.}\right).$$

Taking (40) into account, we then obtain $\mathbb{P}\left(A_{n_k}S_{n_k} > (1-\varepsilon)\chi(n_k)\right.$ i.o.$) = 1$. We thus get $\limsup_{n\to\infty} A_nS_n/\chi(n) \geq 1$ a.s., which proves (32), and concludes the proof of Lemma 5.

3.5.2 Proof of Lemma 4

Now, set

$$\tilde{V}_k = \begin{pmatrix} h_k^{-d/2}\left[\nabla K\left(\frac{\theta-X_k}{h_k}\right) - \mathbb{E}\left(\nabla K\left(\frac{\theta-X_k}{h_k}\right)\right)\right] \\ \tilde{h}_k^{-d/2}\left[K\left(\frac{\theta-X_k}{\tilde{h}_k}\right) - \mathbb{E}\left(K\left(\frac{\theta-X_k}{\tilde{h}_k}\right)\right)\right] \end{pmatrix}$$

and $\Gamma_k = \mathbb{E}(\tilde{V}_k\tilde{V}_k^T)$. In view of (24), (25) and (26), we have $\lim_{k\to\infty}\Gamma_k = \Gamma$. It follows that $\exists k_0 \geq 1$ such that $\forall k \geq k_0$, Γ_k is inversible; without loss of generality, we assume $k_0 = 1$, and set $\tilde{U}_k = \Gamma_k^{-1/2}\tilde{V}_k$. Set $p \in]2,4[$ and let $\mathcal{L}$ be a slowly varying function; we have:

$$\begin{aligned} \frac{\mathbb{E}\big(\|\tilde{U}_k\|^p\big)}{(k\log\log k)^{p/2}} &= O\left(\frac{h_k^{-dp/2}\mathbb{E}\left[\|\nabla K\left(\frac{\theta-X_k}{h_k}\right)\|^p\right] + \tilde{h}_k^{-dp/2}\mathbb{E}\left[\left|K\left(\frac{\theta-X_k}{h_k}\right)\right|^p\right]}{(k\log\log k)^{p/2}}\right) \\ &= O\left(\frac{h_k^{d-dp/2} + \tilde{h}_k^{d-dp/2}}{(k\log\log k)^{p/2}}\right) \\ &= O\left(\mathcal{L}(k)\left[k^{-\left[1+(\frac{p}{2}-1)(1-ad)\right]} + k^{-\left[1+(\frac{p}{2}-1)(1-\tilde{a}d)\right]}\right]\right) \end{aligned}$$

so that $\sum_k (k\log\log k)^{-p/2}\mathbb{E}(\|\tilde{U}_k\|^p) < \infty$. By application of Theorem 2 of Einmahl [7], we deduce that $\sum_{k=1}^n \tilde{U}_k - \sum_{k=1}^n \eta_k = o(\sqrt{n\log\log n})$ a.s., where η_k are independent, and $\mathcal{N}(0, I_{d+1})$-distributed random vectors. It follows that

$$\sum_{k=1}^n \Gamma^{1/2}\Gamma_k^{-1/2}\tilde{V}_k - \sum_{k=1}^n \varepsilon_k = o(\sqrt{n\log\log n}) \quad \text{a.s.} \tag{41}$$

Now,

$$\begin{aligned} &\Delta_n\left[\sum_{k=1}^n Q_k\Gamma^{1/2}\Gamma_k^{-1/2}\tilde{V}_k - \sum_{k=1}^n Q_k\varepsilon_k\right] \\ &= \Delta_n\sum_{k=1}^n Q_k\left[\Gamma^{1/2}\Gamma_k^{-1/2}\tilde{V}_k - \varepsilon_k\right] \\ &= \Delta_n\sum_{k=1}^n Q_k\left(\sum_{j=1}^k\left[\Gamma^{1/2}\Gamma_j^{-1/2}\tilde{V}_j - \varepsilon_j\right] - \sum_{j=1}^{k-1}\left[\Gamma^{1/2}\Gamma_j^{-1/2}\tilde{V}_j - \varepsilon_j\right]\right) \quad (\text{with}\sum_{j=1}^0 = 0) \\ &= \Delta_n\sum_{k=1}^{n-1}(Q_k - Q_{k+1})\left(\sum_{j=1}^k\left(\Gamma^{1/2}\Gamma_j^{-1/2}\tilde{V}_j - \varepsilon_j\right)\right) + \Delta_n Q_n\sum_{j=1}^n\left(\Gamma^{1/2}\Gamma_j^{-1/2}\tilde{V}_j - \varepsilon_j\right) \\ &= \Delta_n\sum_{k=1}^{n-1}(Q_k - Q_{k+1})\left[o\left(\sqrt{k\log\log k}\right)\right] + \Delta_n Q_n\left[o\left(\sqrt{n\log\log n}\right)\right] \quad \text{a.s.} \end{aligned}$$

Moreover,

$$\Delta_n \sum_{k=1}^{n-1} (Q_k - Q_{k+1}) \left[o\left(\sqrt{k \log\log k}\right)\right]$$
$$= \begin{pmatrix} \sqrt{\frac{h_n^{d+2}}{n}} \sum_{k=1}^{n-1} \left(h_k^{-\frac{d+2}{2}} - h_{k+1}^{-\frac{d+2}{2}} \right) o\left(\sqrt{k \log\log k}\right) & 0 \\ 0 & \sqrt{\frac{\tilde{h}_n^d}{n}} \sum_{k=1}^{n-1} \left(\tilde{h}_k^{-\frac{d}{2}} - \tilde{h}_{k+1}^{-\frac{d}{2}} \right) \end{pmatrix}$$
$$= \begin{pmatrix} o\left(\sqrt{h_n^{d+2} \log\log n}\right) \sum_{k=1}^{n-1} \left(h_k^{-\frac{d+2}{2}} - h_{k+1}^{-\frac{d+2}{2}} \right) & 0 \\ 0 & o\left(\sqrt{\tilde{h}_n^d \log\log n}\right) \sum_{k=1}^{n-1} \left(\tilde{h}_k^{-\frac{d}{2}} - \tilde{h}_{k+1}^{-\frac{d}{2}} \right) \end{pmatrix}.$$

Set $\phi(s) = [h(s)]^{-\frac{d+2}{2}}$ and $\tilde{\phi}(s) = \left[\tilde{h}(s)\right]^{-\frac{d}{2}}$, and let $u_k \in [k, k+1]$; since ϕ' and $\tilde{\phi}'$ vary regularly with exponent $(a(d+2)/2 - 1)$ and $(\tilde{a}d/2 - 1)$ respectively, we have

$$\sum_{k=1}^{n-1} \left(h_k^{-\frac{d+2}{2}} - h_{k+1}^{-\frac{d+2}{2}} \right) = O\left(\sum_{k=1}^{n-1} \phi'(u_k) \right) = O\left(\int_1^n \phi'(s) ds \right) = O\left(h_n^{-\frac{d+2}{2}} \right)$$

and

$$\sum_{k=1}^{n-1} \left(\tilde{h}_k^{-\frac{d}{2}} - \tilde{h}_{k+1}^{-\frac{d}{2}} \right) = O\left(\sum_{k=1}^{n-1} \tilde{\phi}'(u_k) \right) = O\left(\int_1^n \tilde{\phi}'(s) ds \right) = O\left(\tilde{h}_n^{-\frac{d}{2}} \right),$$

so that $\Delta_n \sum_{k=1}^{n-1} (Q_k - Q_{k+1}) \left[o\left(\sqrt{k \log\log k}\right)\right] = o\left(\sqrt{\log\log n}\right)$. Since $\Delta_n Q_n \left[o\left(\sqrt{n \log\log n}\right)\right] = o\left(\sqrt{\log\log n}\right)$, we deduce that

$$\frac{\Delta_n \sum_{k=1}^n Q_k \Gamma^{1/2} \Gamma_k^{-1/2} \tilde{V}_k}{\sqrt{2 \log\log n}} - \frac{\Delta_n \sum_{k=1}^n Q_k \varepsilon_k}{\sqrt{2 \log\log n}} = o(1) \quad \text{a.s.}$$

The application of Lemma 5 then ensures that, with probability one, the sequence $(\Delta_n \sum_{k=1}^n Q_k \Gamma^{1/2} \Gamma_k^{-1/2} \tilde{V}_k / \sqrt{2 \log\log n})$ is relatively compact and its limit set is $\mathcal{E} = \{\nu \in \mathbb{R}^{d+1} \text{ such that } \nu^T \Sigma^{-1} \nu \leq 1\}$. Since

$$\frac{\Delta_n \sum_{k=1}^n Q_k \tilde{V}_k}{\sqrt{2 \log\log n}} = \frac{\Delta_n \sum_{k=1}^n Q_k \Gamma^{1/2} \Gamma_k^{-1/2} \tilde{V}_k}{\sqrt{2 \log\log n}} + \frac{\Delta_n \sum_{k=1}^n Q_k \left(I_{d+1} - \Gamma^{1/2} \Gamma_k^{-1/2} \right) \tilde{V}_k}{\sqrt{2 \log\log n}}$$

with $\lim_{k \to \infty} (I_{d+1} - \Gamma^{1/2} \Gamma_k^{-1/2}) = 0$, Lemma 4 follows.

3.6 Proof of Theorems 1 and 2

In view of (15) (and the comment below), Theorem 1 (respectively Theorem 2) is a straightforward consequence of the combination of (20), (21) and (22) (respectively (27), (28) and (29)) together with the following lemma, which establishes that the residual term R_n (defined as in (14)) is negligeable.

Lemma 6 *Let Assumptions (A1)-(A5) hold. If (C2) holds, then* $\lim_{n \to \infty} \tilde{h}_n^{-q} R_n = 0$ *a.s. Otherwise,* $\lim_{n \to \infty} \sqrt{n \tilde{h}_n^d} R_n = 0$ *a.s.*

Proof of Lemma 6 We first note that a Taylor's expansion implies the existence of ζ_n such that $\|\zeta_n - \theta_n\| \leq \|\theta_n - \theta\|$ and

$$\begin{aligned} R_n &= (\theta_n - \theta)^T \nabla \tilde{f}_n(\zeta_n) \\ &= (\theta_n - \theta)^T \left[\nabla \tilde{f}_n(\zeta_n) - \nabla f(\zeta_n) + \nabla f(\zeta_n) - \nabla f(\theta)\right]. \end{aligned}$$

Let $\mathcal{V}$ be a compact set that contains θ; for n large enough, we get

$$\begin{aligned} \|R_n\| &= O\left(\|\theta_n - \theta\| \left[\sup_{x\in\mathcal{V}} \|\nabla \tilde{f}_n(x) - \nabla f(x)\| + \|\zeta_n - \theta\|\right]\right) \\ &= O\left(\|\theta_n - \theta\| \sup_{x\in\mathcal{V}} \|\nabla \tilde{f}_n(x) - \nabla f(x)\| + \|\theta_n - \theta\|^2\right). \end{aligned}$$

On the one hand, let us recall that the a.s. convergence rate of $(\theta_n - \theta)$ is given by the one of $\left[D^2 f(\theta)\right]^{-1} \nabla f_n(\theta)$ (see (15) and the comment below). One can apply (27), (28), and (29) and obtain the exact a.s. convergence rate of $\theta_n - \theta$. However, to avoid assuming (A6), we apply here Lemmas 1 and 2 (with $|\alpha| = 1$ and $(g_n, b_n) = (\tilde{f}_n, \tilde{h}_n)$), and get the following upper bound of the a.s. convergence rate of $\theta_n - \theta$: for any $\gamma > 0$ and $\varepsilon > 0$ small enough,

$$\|\theta_n - \theta\| = O\left(\sqrt{\frac{(\log n)^{1+\gamma}}{n h_n^{d+2}}} + \frac{\sum_{i=1}^n h_i^q}{n}\right) = O\left(\sqrt{\frac{(\log n)^{1+\gamma}}{n h_n^{d+2}}} + h_n^{q-\varepsilon}\right) \quad \text{a.s.} \tag{42}$$

On the other hand, we have

$$\sup_{x\in\mathcal{V}} \|\nabla \tilde{f}_n(x) - \nabla f(x)\| \leq \sup_{x\in\mathcal{V}} \|\nabla \tilde{f}_n(x) - \mathbb{E}\left(\nabla \tilde{f}_n(x)\right)\| + \sup_{x\in\mathcal{V}} \|\mathbb{E}\left(\nabla \tilde{f}_n(x)\right) - \nabla f(x)\|.$$

The application of Lemmas 1 and 2 with $|\alpha| = 1$, $(g_n, b_n) = (\tilde{f}_n, \tilde{h}_n)$ ensures that, for any $\gamma > 0$ and $\varepsilon > 0$ small enough,

$$\sup_{x\in\mathcal{V}} \|\nabla \tilde{f}_n(x) - \nabla f(x)\| = O\left(\sqrt{\frac{(\log n)^{1+\gamma}}{n \tilde{h}_n^{d+2}}} + \frac{\sum_{i=1}^n \tilde{h}_i^q}{n}\right) = O\left(\sqrt{\frac{(\log n)^{1+\gamma}}{n \tilde{h}_n^{d+2}}} + \tilde{h}_n^{q-\varepsilon}\right) \quad \text{a.s.} \tag{43}$$

Let $\mathcal{L}$ denotes a generic slowly varying function that may vary from line to line.

- Let us first assume that (C1) holds. The application of (42) and (43) ensures that for any $\varepsilon > 0$ small enough,

  $$\sqrt{n\tilde{h}_n^d}\|\theta_n - \theta\| \sup_{x\in\mathcal{V}} \|\nabla \tilde{f}_n(x) - \nabla f(x)\| = O\left(\mathcal{L}(n)\left[n^{-\frac{1}{2}(1-a(d+2)-2\tilde{a})} + n^{\tilde{a}-a(q-\varepsilon)}\right]\right) + o(1) \quad \text{a.s.}$$

  Observe that by (C1)i), it is straightforward to see that $2\tilde{a} + a(d+2) < 1$ and $\tilde{a} < a(q-\varepsilon)$ for any $\varepsilon > 0$ small enough, so that $\sqrt{n\tilde{h}_n^d}\|\theta_n - \theta\| \sup_{x\in\mathcal{V}} \|\nabla \tilde{f}_n(x) - \nabla f(x)\| = o(1)$ a.s. Moreover, the application of (42) ensures that

  $$\sqrt{n\tilde{h}_n^d}\|\theta_n - \theta\|^2 = O\left(\mathcal{L}(n)\left[n^{-\frac{1}{2}(1-2a(d+2)+\tilde{a}d)} + n^{\frac{1}{2}(1-\tilde{a}d-4a(q-\varepsilon))}\right]\right) \quad \text{a.s.}$$

  Now, by (C1)ii) we have $2a(d+2) - \tilde{a}d < 1$ and $\tilde{a}d + 4a(q-\varepsilon) > 1$ for any $\varepsilon > 0$ small enough, and thus it follows that $\sqrt{n\tilde{h}_n^d}\|\theta_n - \theta\|^2 = o(1)$ a.s., which ensures the first part of Lemma 6.

- We now assume that (C2) holds. Since $\tilde{a}q \leq q/(d+2q) < 1$, using (9), (42) and (43), we have

$$\begin{aligned}\frac{1}{\tilde{h}_n^q}\|\theta_n - \theta\| \sup_{x\in\mathcal{V}} \|\nabla \tilde{f}_n(x) - \nabla f(x)\| \\ = \quad O\left(\mathcal{L}(n)\left[n^{-1+\frac{a(d+2)}{2}+\frac{\tilde{a}(d+2q+2)}{2}} + n^{-\frac{1}{2}-a(q-\varepsilon)+\frac{\tilde{a}(d+2q+2)}{2}}\right]\right) + o(1) \quad \text{a.s.} \qquad (44)\end{aligned}$$

On the one hand, for any $\varepsilon > 0$ small enough, it is straightforward to see that condition (C2) implies the following inequalities:

$$a(d+2) + \tilde{a}(d+2q+2) < 2 \quad \text{and} \quad \tilde{a}(d+2+2q) < 1 + 2a(q-\varepsilon), \qquad (45)$$

$$\tilde{a}q + a(d+2) < 1 \quad \text{and} \quad \tilde{a}q < 2a(q-\varepsilon). \qquad (46)$$

Therefore, it follows from (44) and (45) that

$$\frac{1}{\tilde{h}_n^q}\|\theta_n - \theta\| \sup_{x\in\mathcal{V}} \|\nabla \tilde{f}_n(x) - \nabla f(x)\| \quad = \quad o(1) \quad \text{a.s.}$$

On the other hand, observe again that by (42) and (46), we have

$$\frac{1}{\tilde{h}_n^q}\|\theta_n - \theta\|^2 \quad = \quad O\left(\mathcal{L}(n)\left[n^{-(1-\tilde{a}q-a(d+2))} + n^{\tilde{a}q-2a(q-\varepsilon)}\right]\right) = o(1) \quad \text{a.s.},$$

which concludes the proof of Lemma 6.

Acknowledgements We deeply thank two anonymous Referees for their helpfull suggestions and comments.

References

[1] Abraham, C., Biau, G. and Cadre, B. (2003), Simple estimation of the mode of a multivariate density. *Canadian J. Statist.* **31**, pp 23-34.

[2] Abraham, C., Biau, G. and Cadre, B. (2004), On the asymptotic properties of a simple estimate of the mode. *ESAIM Prob. and Statist.* **8**, pp 1-11.

[3] Davies, H.I. (1973), Strong consistency of a sequential estimator of a probability density function. *Bull. Math. Statist.* **15**, pp. 49-54.

[4] Devroye, L. (1979), On the pointwise and integral convergence of recursive kernel estimates of probability densities. *Utilitas Math.* **15**, pp. 113-128.

[5] Eddy, W.F. (1980), Optimum kernel estimates of the mode, *Ann. Statist.* **8**, pp. 870-882.

[6] Eddy, W.F. (1982), The asymptotic distributions of kernel estimators of the mode, *Z. Warsch. Verw. Gebiete* **59**, pp. 279-290.

[7] Einmahl, U. (1987), A useful estimate in the multidimensional invariance principle. *Probability theory and related fields* **76**, pp. 81-101.

[8] Feller, W. (1970), An introduction to probability theory and its applications. Second edition Volume II, Wiley.

[9] Hall, P. (1992), Effect of bias estimation on coverage accuracy of bootstrap confidence intervals for a probability density, *Ann. Statist.* **20**, pp. 675-694.

[10] Konakov, V.D. (1973), On asymptotic normality of the sample mode of multivariate distributions, *Theory Probab. Appl.* **18**, pp. 836-842.

[11] Koval, V. (2002), A new law of the iterated logarithm in $\mathbb{R}^d$ with application to matrix-normalized sums of randoms vectors. *Journal of Theoretical Probability* **15**, pp. 249-257.

[12] Menon, V.V., Prasad, B. and Singh, R.S. (1984), Non-parametric recursive estimates of a probability density function and its derivatives. *Journal of Statistical Planning and inference* **9**, pp. 73-82.

[13] Mokkadem, A. and Pelletier, M. (2003), The law of the iterated logarithm for the multivariate kernel mode estimator. *ESAIM: Probab. Statist.* **7**, pp. 1-21.

[14] Mokkadem, A. and Pelletier, M. (2007), A companion for the Kiefer-Wolfowitz-Blum stochastic approximation algorithm, *Ann. Statist.,* **35**, 1749-1772.

[15] Mokkadem, A., Pelletier, M. and Thiam, B. (2006), Large and moderate deviations principles for recursive kernel estimators of a multivariate density and its partial derivatives. *Serdica Math. J.* **32**, pp. 323-354.

[16] Müller H.G. (1989) Adaptive nonparametric peak estimation *Ann. Statist.,* **17**, 1053-1069.

[17] Nadaraya, E.A. (1965), On non-parametric estimates of density functions and regression curves. *Theory Probab. Appl.* **10**, pp. 186-190.

[18] Parzen, E. (1962), On estimation of a probability density function and mode. *Ann. Math. Statist.* **33**, pp. 1065-1076.

[19] Petrov, V.V. (1995), Limit theorems in probability theory, Clarendon Press, Oxford.

[20] Romano, J. (1988), On weak convergence and optimality of kernel density estimates of the mode. *Ann. Statist.* **16**, pp. 629-647.

[21] Rosenblatt, M. (1956), Remarks on some-non-parametric estimates of density function. *Ann. Math. Statist.* **27**, pp. 832-837.

[22] Roussas, G. (1992), Exact rates of almost sure convergence of a recursive kernel estimate of a probability density function: Application to regression and hazard rate estimate. *J. of Nonparam. Statist.* **3**, pp. 171-195.

[23] Rüschendorf, L. (1977), Consistency of estimators for multivariate density functions and for the mode. *Sankhya Ser. A*, **39**, pp. 243-250.

[24] Samanta, M. (1973), Nonparametric estimation of the mode of a multivariate density, *South African Statist. J.* **7**, pp. 109-117.

[25] Tsybakov, A.B. (1990), Recurrent estimation of the mode of a multidimensional distribution, *Problems Inform. Transmission* **26**, 31-37

[26] Van Ryzin, J. (1969), On strong consistency of density estimates. *Ann. Math. Statist.* **40**, pp 1765-1772.

[27] Vieu, P. (1996), A note on density mode estimation, *Statist. Probab. Lett.* **26**, 297-307

[28] Wegman, E.J. and Davies, H.I. (1979), Remarks on some recursive estimators of a probability density. *Ann. Statist.* **7**, pp. 316-327.

[29] Wertz, W. (1985), Sequential and recursive estimators of the probability density. *Statistics* **16**, pp. 277-295.

[30] Wolverton, C.T. and Wagner, T.J. (1969), Asymptotically optimal discriminant functions for pattern classification. *IEEE Trans. Inform. Theory* **15**, pp. 258-265.

[31] Yamato, H. (1971), Sequential estimation of a continuous probability density function and mode. *Bull. Math. Satist.* **14**, pp. 1-12.

[32] Ziegler, K. (2003), On the asymptotic normality of kernel regression estimators of the mode in the random design model. *J. Statist. Plann. Inf.* **115**, pp. 123-144.

[33] Ziegler, K. (2004), Adaptive kernel estimation of the mode in a nonparametric random design regression model. *Probab. Math. Satist.* **24**, pp. 213-235.