Bernoulli **12**(2), 2006, 351–370

Adaptive density estimation using the blockwise Stein method

PHILIPPE RIGOLLET

Laboratoire de Probabilités et Modèles Aléatoires, Université Paris 6, 4 place Jussieu, Case 188, F-75252 Paris, Cedex 05, France. E-mail: rigollet@ccr.jussieu.fr

We study the problem of nonparametric estimation of a probability density of unknown smoothness in $L_2(\mathbb{R})$. Expressing mean integrated squared error (MISE) in the Fourier domain, we show that it is close to mean squared error in the Gaussian sequence model. Then applying a modified version of Stein's blockwise method, we obtain a linear monotone oracle inequality. Two consequences of this oracle inequality are that the proposed estimator is sharp minimax adaptive over a scale of Sobolev classes of densities, and that its MISE is asymptotically smaller than or equal to that of kernel density estimators with any bandwidth provided that the kernel belongs to a large class of functions including many standard kernels.

Keywords: adaptive density estimation; blockwise Stein rule; kernel oracle; monotone oracle; Oracle inequalities

1. Introduction

A Stein weakly geometrically increasing (WGI) blockwise shrinkage estimator, employing a classical Stein blockwise shrinkage together with WGI blocks, has recently been proposed and studied for a filtering problem in Cavalier and Tsybakov (2001) and Tsybakov (2002). It has been established that the estimator possesses several very nice statistical properties. This paper suggests a Stein WGI estimator for the problem of probability density estimation and then studies its properties via an oracle inequality. It also shows how to use an oracle inequality to obtain Stone type results for kernel estimates.

Consider independent and identically distributed random variables $X_1, \ldots, X_n$ having an unknown common probability density $p \in L_2(\mathbb{R})$. We study the estimation of p based on the sample $\mathbb{X}^n = (X_1, \ldots, X_n)$. Let $\hat{p}_n$ be an estimator of p. We measure the performance of $\hat{p}_n$ by its *mean integrated squared error* (MISE),

$$\mathbb{E}_p\|\hat{p}_n - p\|^2 = \mathbb{E}_p\int_{\mathbb{R}}(\hat{p}_n(x) - p(x))^2\,\mathrm{d}x, \tag{1.1}$$

where $\mathbb{E}_p$ denotes the expectation with respect to $\mathbb{X}^n$. Define the characteristic function $\varphi(t) = \int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}tx}p(x)\mathrm{d}x$ and the empirical characteristic function $\varphi_n(t) = \varphi_n(t, \mathbb{X}^n) = n^{-1}\sum_{k=1}^n \mathrm{e}^{\mathrm{i}tX_k}$. For any function $h \in L_2(\mathbb{R})$, let $\omega \mapsto \mathcal{F}[h](\omega) = \int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}\omega x}h(x)\mathrm{d}x$ be its *Fourier transform* (the integral is understood in the 'limit in mean' sense). Consider a linear estimator of the characteristic function φ defined by

1350–7265

$$\hat{\varphi}_\lambda(\omega) \triangleq \varphi_n(\omega)\lambda(\omega), \tag{1.2}$$

where $\omega \mapsto \lambda(\omega)$ is a *weight function* in $L_2(\mathbb{R})$. We define a density estimator $\hat{p}_\lambda$ of p as the inverse Fourier transform of $\hat{\varphi}_\lambda$. The performance of this new estimator is measured by its MISE which, by the Plancherel equality, can be written as

$$\mathbb{E}_p\|\hat{p}_\lambda - p\|^2 = \frac{1}{2\pi}\mathbb{E}_p\int_{\mathbb{R}} |\hat{\varphi}_\lambda(\omega) - \varphi(\omega)|^2\,\mathrm{d}\omega \triangleq \frac{1}{2\pi}R_n(\hat{\varphi}_\lambda, \varphi),$$

where

$$R_n(\hat{\varphi}_\lambda, \varphi) = \int_{\mathbb{R}}\left(|1-\lambda(\omega)|^2|\varphi(\omega)|^2 + \frac{1}{n}|\lambda(\omega)|^2\right)\mathrm{d}\omega - \frac{1}{n}\int_{\mathbb{R}}|\varphi(\omega)|^2|\lambda(\omega)|^2\,\mathrm{d}\omega.$$

Hence, the MISE of the linear estimator equals the Δ-*risk*

$$\Delta_n\big(\lambda, |\varphi|^2\big) \triangleq \int_{\mathbb{R}}\left(|1-\lambda(\omega)|^2|\varphi(\omega)|^2 + \frac{|\lambda(\omega)|^2}{n}\right)\mathrm{d}\omega, \tag{1.3}$$

minus a residual term of order n^{-1} which is usually small compared to Δ_n. This suggests that, for sufficiently large n, the linear oracle $\lambda_{\mathcal{H}}^{\text{oracle}}$ on a class $\mathcal{H}$ satisfies

$$\lambda_{\mathcal{H}}^{\text{oracle}} \triangleq \arg\min_{\lambda\in\mathcal{H}} R_n(\hat{\varphi}_\lambda, \varphi) \approx \arg\min_{\lambda\in\mathcal{H}}\big[\Delta_n\big(\lambda, |\varphi|^2\big)\big]. \tag{1.4}$$

Note that expression (1.3) is similar to that for the mean squared error in the Gaussian sequence model, transposing the formula in the continuous case and replacing ε by $n^{-1/2}$. Therefore, it seems natural to extend the results for the Gaussian sequence model to nonparametric density estimation. The similarity between density estimation and the Gaussian sequence model based on Fourier analysis has been examined by Golubev (1992) and later by Boiko and Golubev (2000). Golubev and Levit (1996) consider the problem of the second-order minimax adaptive estimation of an unknown distribution function over Sobolev ellipsoids. They develop techniques that are also useful for the density estimation problem considered here.

The Gaussian sequence model has been studied by many authors in the past decades and oracle inequalities have been widely used, although initially in an implicit form, to derive adaptation – see Shibata (1981), Efroimovich and Pinsker (1984), Li (1987), Golubev (1990, 1992), Golubev and Nussbaum (1992), Polyak and Tsybakov (1992), Kneip (1994), Birgé and Massart (2001) and Cavalier *et al.* (2002). Most of these papers use the Mallows C_p or its modifications to derive estimators that mimic the best estimator in various subclasses of linear estimators (i.e. the oracle). In particular, blockwise constant weights show particularly good statistical properties and have been widely discussed in the statistical literature, first by Efroimovich and Pinsker (1984), and more recently by Efromovich (1999), Nemirovski (2000) and Efromovich (2004a) who consider block estimators different from Stein's one. Tsybakov (2004) considers Stein's estimator with a particular system of blocks, namely, WGI blocks. Cavalier and Tsybakov (2001) improve the previous results by using a penalized version of the block Stein estimator.

As for the wavelet framework, the subject has been discussed by Donoho and Johnstone (1994, 1995) and Härdle *et al.* (1998). In the same setting, Cai (1999) and Efromovich (2000) use block thresholding type estimators that both satisfy oracle inequalities within the class of blockwise linear estimators. These estimators exhibit good performance in simulations.

Goldenshluger and Tsybakov (2001) apply the block Stein estimator with WGI blocks to the Gaussian regression problem with infinitely many parameters. They show that it is sharp minimax adaptive over a scale of Sobolev ellipsoids in ℓ_2. Tsybakov (2002) discusses in particular the anisotropic multidimensional white noise model. He shows that the block Stein estimator, again with WGI blocks, is adaptive simultaneously with respect to the real dimension, direction and smoothness of the parameter over a scale of Sobolev ellipsoids. In both papers, adaptation is derived from oracle inequalities.

Whereas the Gaussian sequence model has been extensively studied, there are few results concerning blockwise density estimation. A discussion of blockwise density estimates can be found in Efroimovich (1985) and Efromovich (2000, 2005), where the Efromovich–Pinsker shrinkage procedure together with polynomial blocks is explored, and in Hall *et al.* (1998), where a block-thresholding shrinkage procedure employing small logarithmic blocks is explored.

Cavalier and Tsybakov (2001) obtained powerful oracle inequalities for penalized Stein estimates in the context of the Gaussian sequence model. Under certain hypotheses, they lead to adaptive properties in the minimax sense – in particular, over any ellipsoid in ℓ_2 with monotone decreasing coefficients.

This paper is devoted to developing a Stein WGI estimator for a density estimation setting. The estimator employs a classical Stein blockwise shrinkage which uses a zero thresholding (imposes no penalty). Let us recall that this shrinkage procedure has been very attractive for filtering problems; see the discussion in Tsybakov (2002). The WGI blocks employed were also recommended in Tsybakov (2002). Note that these blocks are larger than the ones traditionally studied in the literature, but this choice is justified in Efromovich (2004b) where it is shown that the smaller blocks do not imply sharp minimaxity.

The primary complication in the development and study of a density estimate, based on a known analogue for Gaussian sequence models, consists in the fact that the observations are not Gaussian and precise results such as Stein's lemma do not apply. In Section 2, we use unbiased estimation of the risk to derive several oracle inequalities for the proposed estimator. We then give two corollaries of these results. First, in Section 3 we show that it is sharp minimax adaptive over a scale of Sobolev classes of densities. Second, in Section 4 we show that its MISE is asymptotically smaller than or equal to that of kernel density estimators with any bandwidth provided that the kernel belongs to a large class of functions including many standard kernels. A simulation study for the proposed estimator can be found in Rigollet (2004).

2. Application of the blockwise Stein method to density estimation

2.1. Estimation of the Δ-risk

Blockwise Stein methods in the Gaussian sequence model are related to the *unbiased estimation of the risk*. For density estimation, we will consider only an asymptotically unbiased estimator of the risk. The empirical characteristic function φ_n is an unbiased estimator of φ that satisfies

$$\mathbb{E}_p|\varphi_n(t)|^2 = \left(1 - \frac{1}{n}\right)|\varphi(t)|^2 + \frac{1}{n}. \tag{2.1}$$

Therefore, $[|\varphi_n(\omega)|^2 - 1/n]$ is an asymptotically unbiased estimator of $|\varphi(\omega)|^2$ and we can define an asymptotically unbiased estimator of $\Delta_n(\lambda, |\varphi|^2)$ which, up to an additional term independent of λ, is given by

$$l_n(\lambda) \triangleq \int_{\mathbb{R}} \left(\left[|\lambda(\omega)|^2 - 2\mathrm{Re}(\lambda(\omega)) \right] |\varphi_n(\omega)|^2 + \frac{2}{n}\mathrm{Re}[\lambda(\omega)] \right) \mathrm{d}\omega, \tag{2.2}$$

for λ such that the integrals are finite. It is natural to drop this additional term since the idea is to choose a weight function λ that minimizes l_n over a certain class $\mathcal{H}$. To define a reasonable $\mathcal{H}$, we shall restrict the possible weight functions λ to *admissible* ones. Cline (1988) proves that if λ is an arbitrary complex-valued function, it is possible to find a real, non-negative function $\bar{\lambda}$, bounded by one, such that the risk corresponding to $\bar{\lambda}$ is smaller than that corresponding to λ. All such $\bar{\lambda}$ will be called *admissible*. For all admissible λ, the functional $l_n(\lambda)$ defined in (2.2) becomes

$$l_n(\lambda) = \int_{\mathbb{R}} \left(\left[\lambda^2(\omega) - 2\lambda(\omega) \right] |\varphi_n(\omega)|^2 + \frac{2}{n}\lambda(\omega) \right) \mathrm{d}\omega. \tag{2.3}$$

We also impose the restriction that λ is even. From now on, let $\mathcal{H}_0$ be the class of all even, square-integrable functions on $\mathbb{R}$ taking values in [0, 1].

We first study a simple class of weight functions which leads to the definition of an analogue of the Stein estimator for density estimation. For this estimator, a first oracle inequality is given. Then the construction is generalized to a slightly more complex class of weight functions to define the blockwise Stein estimator.

2.2. Stein's estimators applied to density estimation

Consider a particularly simple class of weight functions,

$$\mathcal{H}_A \triangleq \{\lambda : \lambda(\omega) = t\mathbb{1}_A(\omega),\ t \in [0, 1]\} \subset \mathcal{H}_0,$$

where A is a finite union of bounded, non-trivial intervals on $\mathbb{R}$ such that $a \in A \Rightarrow -a \in A$ (later we use the union of two intervals symmetric about 0, which we will call *symmetrized* *intervals*) and $\mathbb{1}_A$ denotes the indicator function of A. Define the *Stein estimator on A* as the solution of the minimization problem $\lambda_A^* \triangleq \arg\min_{\lambda\in\mathcal{H}_A} l_n(\lambda)$, which can explicitly be written as

$$\lambda_A^*(\omega) = \left(1 - \frac{|A|}{n\int_A |\varphi_n|^2}\right)_+ \mathbb{1}_A(\omega) \triangleq t_A^* \mathbb{1}_A(\omega), \qquad \omega \in \mathbb{R}, \tag{2.4}$$

where $|A|$ is the Lebesgue measure of A. Then let $\lambda_A^{\text{oracle}}$ be the *linear oracle on* $\mathcal{H}_A$, defined by

$$\lambda_A^{\text{oracle}} \triangleq \arg\min_{\lambda\in\mathcal{H}_A} R_n^A(\hat{\varphi}_\lambda, \varphi), \qquad \text{where } R_n^A(\hat{\varphi}_\lambda, \varphi) \triangleq \mathbb{E}_p \int_A \left|\hat{\varphi}_\lambda(\omega) - \varphi(\omega)\right|^2 \mathrm{d}\omega. \tag{2.5}$$

It is easy to see that

$$\lambda_A^{\text{oracle}}(\omega) = \left(\frac{\int_A |\varphi|^2}{\int_A |\varphi|^2 + n^{-1}\int_A (1 - |\varphi|^2)}\right) \mathbb{1}_A(\omega) \triangleq t_A^{\text{oracle}} \mathbb{1}_A(\omega). \tag{2.6}$$

Theorem 2.1. *Let $1 \leqslant |A| \leqslant 4n$ and let φ satisfy $\int_A |\varphi(\omega)|\mathrm{d}\omega \leqslant G$, for some $G < \infty$. Then there exist an absolute constant $C > 0$ and a constant $D_1 > 0$ that depends only on G such that, for any $\mu_n > C$, the Stein estimator on the set A satisfies the oracle inequality*

$$R_n^A\left(\hat{\varphi}_{\lambda_A^*}, \varphi\right) \leqslant \frac{1}{1 - C\mu_n^{-1}} \left(R_n^A\left(\hat{\varphi}_{\lambda_A^{\text{oracle}}}, \varphi\right) + D_1 \frac{(\log n)^4 \mu_n}{n}\right). \tag{2.7}$$

The proof of Theorem 2.1 is given in Section 6.

Now introduce a constant $b_0 > 0$ and a finite value Ω_n depending only on n and consider a partition $\{B_j\}_{j=0}^J$ of $[-\Omega_n, \Omega_n]$, such that $B_0 = (-b_0, b_0)$ and,

$$\forall 1 \leqslant j \leqslant J, \qquad B_j = -B_j' \cup B_j', \quad B_j' \triangleq [b_{j-1}, b_j), \quad -B_j' \triangleq (-b_j, -b_{j-1}], \quad 0 < b_{j-1} < b_j.$$

Let $\mathcal{H}^*$ be the class of weight functions given by

$$\mathcal{H}^* = \left\{\lambda : \lambda(\omega) = \sum_{j=0}^J t_j \mathbb{1}_{B_j}(\omega),\, 0 \leqslant t_j \leqslant 1,\, j = 0, \ldots, J\right\} \subset \mathcal{H}_0. \tag{2.8}$$

Minimization of l_n over $\mathcal{H}^*$ follows directly from the minimization over $\mathcal{H}_{B_j}$. Indeed, the function $\tilde{\lambda} \triangleq \arg\min_{\lambda\in\mathcal{H}^*} l_n(\lambda)$ is constant on each B_j, $\tilde{\lambda}(\omega) = \sum_{j=0}^J \lambda_{B_j}^* \mathbb{1}_{B_j}(\omega)$, where $\lambda_{B_j}^*$ is defined in (2.4). Define the *blockwise Stein estimator on the system* $\{B_j\}_{j=0}^J$ by

$$\tilde{\lambda}(\omega) \triangleq \sum_{j=0}^J \left(1 - \frac{|B_j|}{n\int_{B_j} |\varphi_n|^2}\right)_+ \mathbb{1}_{B_j}(\omega). \tag{2.9}$$

From (2.7) we obtain an oracle inequality for the blockwise Stein estimator. Indeed, if every B_j, $j = 0, \ldots, J$, satisfies $1 \leqslant |B_j| \leqslant 4n$, by Theorem 2.1, for $\mu_n > C$ and $\int |\varphi| \leqslant G$ we have

$$R_n(\hat{\varphi}_{\tilde{\lambda}}, \varphi) = \sum_{j=0}^{J} R_n^{B_j}\left(\hat{\varphi}_{\lambda^*_{B_j}}, \varphi\right) + \int_{|\omega|>\Omega_n} |\varphi(\omega)|^2 \,\mathrm{d}\omega$$

$$\leqslant \frac{1}{1 - C\mu_n^{-1}}\left(\sum_{j=0}^{J} R_n^{B_j}\left(\hat{\varphi}_{\lambda^{\text{oracle}}_{B_j}}, \varphi\right) + JD_1 \frac{(\log n)^4 \mu_n}{n}\right) + \int_{|\omega|>\Omega_n} |\varphi(\omega)|^2 \,\mathrm{d}\omega. \quad (2.10)$$

Let $\lambda^{\text{oracle}}_{\mathcal{H}^*}$ be the linear blockwise constant oracle defined by $\lambda^{\text{oracle}}_{\mathcal{H}^*} \triangleq \arg\min_{\lambda \in \mathcal{H}^*} R_n(\hat{\varphi}_\lambda, \varphi)$. Since

$$\sum_{j=0}^{J} R_n^{B_j}\left(\hat{\varphi}_{\lambda^{\text{oracle}}_{B_j}}, \varphi\right) + \int_{|\omega|>\Omega_n} |\varphi(\omega)|^2 \,\mathrm{d}\omega = R_n(\hat{\varphi}_{\lambda^{\text{oracle}}_{\mathcal{H}^*}}, \varphi),$$

equation (2.10) implies the following result.

Theorem 2.2. *Let $1 \leqslant |B_j| \leqslant 4n$, for any $j = 0, \ldots, J$, and let φ satisfy $\int_{B_j} |\varphi(\omega)| \mathrm{d}\omega \leqslant G$, for any $j = 0, \ldots, J$ and some $G < \infty$. Then there exist an absolute constant $C > 0$ and a constant $D_1 > 0$ that depends only on G such that, for any $\mu_n > C$, the blockwise Stein estimator on the system $\{B_j\}_{j=0}^J$ satisfies the oracle inequality*

$$R_n(\hat{\varphi}_{\tilde{\lambda}}, \varphi) \leqslant \frac{1}{1 - C\mu_n^{-1}}\left(R_n(\hat{\varphi}_{\lambda^{\text{oracle}}_{\mathcal{H}^*}}, \varphi) + JD_1 \frac{(\log n)^4 \mu_n}{n}\right). \quad (2.11)$$

2.3. Linear monotone oracle

Consider the classes of 'monotone' weight functions

$$\mathcal{H}^\infty_{\text{mon}} \triangleq \{\lambda \in \mathcal{H}_0 : \lambda(\omega) \leqslant \lambda(\omega'),\ 0 \leqslant \omega' \leqslant \omega\} \quad \text{and} \quad \mathcal{H}^{\Omega_n}_{\text{mon}} \triangleq \left\{\lambda \mathbb{1}_{[-\Omega_n, \Omega_n]} : \lambda \in \mathcal{H}^\infty_{\text{mon}}\right\}.$$

The space $L_2(\mathbb{R})$ equipped with the $\|\cdot\|$-norm is a reflexive Banach space, and $\mathcal{H}^{\Omega_n}_{\text{mon}}$ is a closed convex subset of $L_2(\mathbb{R})$. Moreover, the functional $F : \lambda \mapsto R_n(\hat{\varphi}_\lambda, \varphi)$ is quadratic and coercive (i.e. $F(\lambda) \to +\infty$, $\|\lambda\| \to \infty$) and thus strictly convex and continuous. These remarks and Proposition 1.2 of Ekeland and Temam (1974, Chapter II) show that there exists a unique $\lambda^{\Omega_n}_{\text{mon}} \triangleq \arg\min_{\lambda \in \mathcal{H}^{\Omega_n}_{\text{mon}}} R_n(\hat{\varphi}_\lambda, \varphi)$. Similarly there exists a unique $\lambda^\infty_{\text{mon}} \triangleq \arg\min_{\lambda \in \mathcal{H}^\infty_{\text{mon}}} R_n(\hat{\varphi}_\lambda, \varphi)$. We call $\lambda^{\Omega_n}_{\text{mon}}$ and $\lambda^\infty_{\text{mon}}$ *linear monotone oracles*. In the Gaussian sequence model, under some assumptions on the system of blocks, the blockwise Stein estimator is almost as good as the linear monotone oracle (Tsybakov 2004). An analogue of this result for density estimation is given below. Let the system of symmetrized intervals satisfy the following assumption.

Assumption A. *The inequality*

$$\max_{0 \leqslant j \leqslant J-1} \frac{|B_{j+1}|}{|B_j|} \leqslant 1 + \eta,$$

holds for some $\eta > 0$.

Lemma 2.1. *Under Assumption A, for all* $\varphi \in L_2(\mathbb{R})$,

$$\begin{aligned}\min_{\lambda\in\mathcal{H}^*} R_n(\hat{\varphi}_\lambda, \varphi) &\leqslant \min_{\lambda\in\mathcal{H}^*\cap\mathcal{H}^{\Omega_n}_{\mathrm{mon}}} R_n(\hat{\varphi}_\lambda, \varphi)\\ &\leqslant (1+\eta)\min_{\lambda\in\mathcal{H}^{\Omega_n}_{\mathrm{mon}}} R_n(\hat{\varphi}_\lambda, \varphi) + \frac{1}{n}\big(|B_0| + 3(1+\eta)\|\varphi\|^2\big).\end{aligned} \tag{2.12}$$

Proof. We need to show that for any $\lambda \in \mathcal{H}^{\Omega_n}_{\mathrm{mon}}$, there exists $\bar{\lambda} \in \mathcal{H}^* \cap \mathcal{H}^{\Omega_n}_{\mathrm{mon}}$ such that

$$R_n(\hat{\varphi}_{\bar{\lambda}}, \varphi) \leqslant (1+\eta)R_n(\hat{\varphi}_\lambda, \varphi) + \frac{1}{n}\big(|B_0| + 3(1+\eta)\|\varphi\|^2\big). \tag{2.13}$$

Fix $\lambda \in \mathcal{H}^{\Omega_n}_{\mathrm{mon}}$ and define $\tilde{\lambda}(\omega) \triangleq \min\Big[\lambda(\omega), \big(1 + n^{-1/2}\big)^{-1}\Big]$. Inequality (2.13) holds for $\bar{\lambda}(\omega) \triangleq \sum_{j=0}^{J}\bar{\lambda}_{(j)}\mathbb{1}_{\{\omega\in B_j\}}$, where $\bar{\lambda}_{(j)} \triangleq \sup_{f\in B_j}\tilde{\lambda}(f)$. Indeed,

$$R_n(\hat{\varphi}_{\bar{\lambda}}, \varphi) \leqslant \int_{\mathbb{R}}\left((1-\tilde{\lambda}(\omega))^2|\varphi(\omega)|^2 + \frac{1}{n}\bar{\lambda}^2(\omega)\right)\mathrm{d}\omega - \frac{1}{n}\int_{\mathbb{R}}|\varphi(\omega)|^2\tilde{\lambda}^2(\omega)\mathrm{d}\omega.$$

But $\bar{\lambda}$ satisfies

$$\int_{\mathbb{R}}\bar{\lambda}^2(\omega)\mathrm{d}\omega = \int_{-\Omega_n}^{\Omega_n}\bar{\lambda}^2(\omega)\mathrm{d}\omega \leqslant |B_0| + (1+\eta)\int_{\mathbb{R}}\tilde{\lambda}^2(\omega)\mathrm{d}\omega.$$

Since

$$\int_{\mathbb{R}}\left((1-\tilde{\lambda}(\omega))^2 - \frac{\tilde{\lambda}^2(\omega)}{n}\right)|\varphi(\omega)|^2\,\mathrm{d}\omega \geqslant 0,$$

it follows that

$$R_n(\hat{\varphi}_{\bar{\lambda}}, \varphi) \leqslant (1+\eta)R_n(\hat{\varphi}_{\tilde{\lambda}}, \varphi) + \frac{|B_0|}{n}. \tag{2.14}$$

On the other hand,

$$R_n(\hat{\varphi}_{\tilde{\lambda}}, \varphi) \leqslant \int_{\mathbb{R}}(1-\tilde{\lambda}(\omega))^2|\varphi(\omega)|^2\,\mathrm{d}\omega + \frac{1}{n}\int_{\mathbb{R}}\big(1 - |\varphi(\omega)|^2\big)\lambda^2(\omega)\mathrm{d}\omega.$$

But, if we note $\int f \triangleq \int_{\mathbb{R}} f(\omega)\mathrm{d}\omega$ for a function f, the first term of the right-hand side becomes

$$\begin{aligned}\int(1-\tilde{\lambda})^2|\varphi|^2 &= \int(1-\lambda)^2|\varphi|^2 + \int(\lambda-\tilde{\lambda})^2|\varphi|^2 + 2\int(1-\lambda)(\lambda-\tilde{\lambda})|\varphi|^2\\ &\leqslant \int(1-\lambda)^2|\varphi|^2 + 3\frac{\|\varphi\|^2}{(1+\sqrt{n})^2}.\end{aligned}$$

Therefore, $R_n(\hat{\varphi}_{\tilde{\lambda}}, \varphi) \leqslant R_n(\hat{\varphi}_\lambda, \varphi) + 3\|\varphi\|^2/n$, which, combined with (2.14), proves the lemma. □

Theorem 2.3. *Let the system $\{B_j\}_{j=0}^J$ satisfy Assumption A and let the conditions of Theorem 2.2 hold. Then there exist an absolute constant $C > 0$ and a constant $D_1 > 0$ that depends only on G such that, for any $\mu_n > C$, the blockwise Stein estimator on the system $\{B_j\}_{j=0}^J$ satisfies the oracle inequality*

$$R_n(\hat{\varphi}_{\tilde{\lambda}}, \varphi) \leqslant \frac{1}{1 - C\mu_n^{-1}} \left((1+\eta) R_n(\hat{\varphi}_{\lambda_{\mathrm{mon}}^{\Omega_n}}, \varphi) + \frac{1}{n} \left(|B_0| + 3(1+\eta)G + JD_1 (\log n)^4 \mu_n \right) \right). \tag{2.15}$$

Proof. The proof follows directly from Theorem 2.2 and Lemma 2.1. □

The next lemma allows us to extend the oracle inequality (2.15) to the class $\mathcal{H}_{\mathrm{mon}}^{\infty}$ of monotone weight functions that do not necessarily have a compact support. Set $\Omega_n = n(\log n)^2$ so that, for sufficiently large n, $\Omega_n \geqslant Gn \log n$.

Lemma 2.2. *Assume that $\|\varphi\|^2 \leqslant G$. For $n \geqslant n_0(G) > 0$, there exist positive constants $\kappa_1 = \kappa_1(n_0)$ and $\kappa_2 = \kappa_2(n_0)$ such that*

$$\min_{\lambda \in \mathcal{H}_{\mathrm{mon}}^{\Omega_n}} R_n(\hat{\varphi}_\lambda, \varphi) \leqslant \left(1 + \frac{\kappa_1}{\log n}\right) \min_{\lambda \in \mathcal{H}_{\mathrm{mon}}^{\infty}} R_n(\hat{\varphi}_\lambda, \varphi) + \kappa_2 \frac{G}{n}.$$

Proof. Define

$$\Omega_n^0 \triangleq \max\left\{ |\omega| : |\lambda(\omega)| \geqslant (\log n)^{-1/2} \right\}.$$

Then, setting $\lambda \triangleq \lambda_{\mathrm{mon}}^{\infty}$ and $\lambda_0 \equiv 0 \in \mathcal{H}_{\mathrm{mon}}^{\infty}$,

$$\|\varphi\|^2 = R_n(\hat{\varphi}_{\lambda_0}, \varphi) \geqslant R_n(\hat{\varphi}_\lambda, \varphi) \geqslant \frac{1}{n} \int_{|\omega| \leqslant \Omega_n^0} \frac{1}{\log n} \mathrm{d}\omega - \frac{\|\varphi\|^2}{n} = \frac{2\Omega_n^0}{n \log n} - \frac{\|\varphi\|^2}{n}.$$

Thus $\Omega_n^0 \leqslant \|\varphi\|^2 n \log n \leqslant \Omega_n$. Now define $\lambda_{\Omega_n}(\omega) \triangleq \lambda(\omega) \mathbb{1}_{\{|\omega| \leqslant \Omega_n\}}$; therefore $\lambda_{\Omega_n} \in \mathcal{H}_{\mathrm{mon}}^{\Omega_n}$ and

$$\begin{aligned} R_n(\hat{\varphi}_{\lambda_{\Omega_n}}, \varphi) &\leqslant \int_{\mathbb{R}} \left[(1 - \lambda(\omega))^2 |\varphi(\omega)|^2 + \frac{\lambda_{\Omega_n}^2(\omega)}{n} \right] \mathrm{d}\omega \\ &\leqslant \left(1 - \frac{1}{\sqrt{\log n}}\right)^{-2} \int_{\mathbb{R}} \left[(1 - \lambda(\omega))^2 |\varphi(\omega)|^2 + \frac{\lambda(\omega)^2}{n} \right] \mathrm{d}\omega \\ &\leqslant \left(1 + \frac{\kappa_1}{\log n}\right) R_n(\hat{\varphi}_\lambda, \varphi) + \kappa_2 \frac{G}{n}, \qquad \text{for } n \geqslant n_0. \end{aligned}$$

□

An important question is how to construct systems of symmetrized intervals $\{B_j\}_{j=0}^J$ satisfying the assumptions of Theorem 2.3 and such that the residual term on the right-hand side of inequality (2.15) is asymptotically negligible with respect to the principal term $(1+\eta)R_n(\hat{\varphi}_{\lambda_{\mathrm{mon}}^{\Omega_n}}, \varphi)$ under rather general conditions on φ. We now give an example of such a construction. In what follows, set $\Omega_n = n^\alpha (\log n)^{\alpha'}$, where $\alpha \geqslant 1/2$, $\alpha' \geqslant 0$. Let ν_n be a deterministic quantity such that $\nu_n \to \infty$ when $n \to \infty$.

Set $\eta_n \triangleq 1/\nu_n$ and define the system of symmetrized intervals $\{B_j\}_{j=0}^J$ with the size $|B_j|$ of each symmetrized interval B_j:

$$
\begin{aligned}
|B_0| &= \nu_n, \\
|B_j| &= (1+\eta_n)^j \nu_n, \qquad j = 1, 2, \ldots, J-1, \\
|B_J| &= \Omega_n - \sum_{j=0}^{J-1} |B_j|,
\end{aligned}
$$

where

$$
J \triangleq \min\left\{ m : \sum_{j=0}^{m} (1+\eta_n)^j \nu_n \geqslant \Omega_n \right\}.
$$

Clearly, this system satisfies Assumption A with $\eta = \eta_n$. We call the system $\{B_j\}_{j=0}^J$ *a weakly geometrically increasing system of symmetrized intervals* or *WGI system*. The corresponding blockwise Stein estimator is called *the Stein WGI estimator*. For $n \geqslant 2$,

$$
\sum_{j=0}^{J-1} (1+\eta_n)^j \nu_n \leqslant \Omega_n. \tag{2.16}
$$

Solving inequality (2.16) with respect to J, we find that there exist $n_0 \geqslant 2$ and $C = C(n_0, \alpha, \alpha')$ such that $J \leqslant C(\log n)\nu_n$, for $n \geqslant n_0$. For the Stein WGI estimator, by the Plancherel identity, inequality (2.15) yields the following theorem.

Theorem 2.4. *Assume that* $\int_{\mathbb{R}} |\varphi| \leqslant G$. *Then there exist an absolute constant* $C > 0$ *and a constant* $D_2 > 0$ *that depends only on G such that, for any* $\mu_n > C$ *and sufficiently large n, the Stein WGI estimator satisfies the oracle inequality*

$$
R_n(\hat{p}_\lambda, p) \leqslant \frac{1}{1 - C\mu_n^{-1}} \left((1+\nu_n^{-1}) R_n(\hat{p}_{\lambda_{\mathrm{mon}}^{\Omega_n}}, p) + \frac{\tau_n(D_2)}{2\pi} \right), \tag{2.17}
$$

where the residual

$$
\tau_n(D_2) \triangleq \frac{D_2 \nu_n}{n} \left(1 + (\log n)^5 \mu_n \right).
$$

3. Application to sharp minimax adaptation

For any $Q>0$ and $\beta>0$, define the *Sobolev class* of densities $\Theta(\beta, Q)$ as the set of functions

$$\Theta(\beta, Q) \triangleq \left\{ p \in L_2(\mathbb{R}),\ p \geqslant 0,\ \int_{\mathbb{R}} p(x)\mathrm{d}x = 1,\ \int_{\mathbb{R}} |\varphi(\omega)|^2 |\omega|^{2\beta}\,\mathrm{d}\omega \leqslant 2\pi Q \right\}.$$

We will show that the Stein WGI estimator is sharp minimax adaptive on a wide scale of Sobolev classes, that is, it is asymptotically sharp minimax simultaneously on all the classes $\Theta(\beta, Q)$, for $\beta > 1/2$ and $Q>0$. Set $\beta>1/2$ and $Q>0$ and define the *Pinsker type density estimator* which is a linear estimator of the form (1.2) with the weight function

$$\ell(\omega) \triangleq \left(1-\kappa^*|\omega|^\beta\right)_+, \qquad \text{where } \kappa^* \triangleq \left(\frac{\beta}{(2\beta+1)(\beta+1)\pi Q}\right)^{\beta/(2\beta+1)} n^{-\beta/(2\beta+1)}. \tag{3.1}$$

It is obvious that, for sufficiently large n, $\ell \in \mathcal{H}_{\mathrm{mon}}^{\sqrt{n}}$. Now, if $p \in \Theta(\beta, Q)$, for $\beta>1/2$ we have $\int_{\mathbb{R}} |\varphi| \leqslant C(\beta, Q)$, where $C(\beta, Q) < \infty$ is a positive constant that depends only on β and Q. Taking the supremum over a Sobolev class of densities $\Theta(\beta, Q)$ of both sides of (2.17), we obtain

$$\sup_{p\in\Theta(\beta,Q)} R_n(\hat{p}_{\tilde{\lambda}}, p) \leqslant \frac{1}{1-C\mu_n^{-1}} \left((1+\nu_n^{-1}) \sup_{p\in\Theta(\beta,Q)} R_n(\hat{p}_{\lambda_{\mathrm{mon}}^{\Omega_n}}, p) + \frac{\tau_n(D(Q,\beta))}{2\pi} \right), \tag{3.2}$$

where $D(Q, \beta) < \infty$ depends only on Q and β. But $\ell \in \mathcal{H}_{\mathrm{mon}}^{\sqrt{n}}$, so for every $p \in \Theta(\beta, Q)$ with Fourier transform φ, we have that $R_n(\hat{\varphi}_{\lambda_{\mathrm{mon}}^{\sqrt{n}}}, \varphi) \leqslant R_n(\hat{\varphi}_\ell, \varphi)$, which implies that $R_n(\hat{p}_{\lambda_{\mathrm{mon}}^{\sqrt{n}}}, p) \leqslant R_n(\hat{p}_\ell, p)$. For $\Omega_n = \sqrt{n}$, inequality (3.2) with, for instance, $\mu_n = \nu_n = \log(\log n)$ yields

$$\begin{aligned} \sup_{p\in\Theta(\beta,Q)} R_n(\hat{p}_{\tilde{\lambda}}, p) &\leqslant \frac{1}{1-C\mu_n^{-1}} \left((1+\nu_n^{-1}) \sup_{p\in\Theta(\beta,Q)} R_n(\hat{p}_\ell, p) + \frac{\tau_n(D(Q,\beta))}{2\pi} \right) \\ &\leqslant C^* n^{-2\beta/(2\beta+1)}(1+o(1)), \qquad n \to +\infty, \end{aligned} \tag{3.3}$$

where

$$C^* \triangleq (2\beta+1)\left[\frac{\pi(2\beta+1)(\beta+1)}{\beta}\right]^{-2\beta/(2\beta+1)} Q^{1/(2\beta+1)}.$$

The last inequality in (3.3) is a known upper bound for the Pinsker type density estimator that is proved by Schipper (1996) for integer β, although the proof remains valid for any positive β. On the other hand, the following lower bound holds.

$$\inf_{T_n} \sup_{p\in\Theta(\beta,Q)} n^{2\beta/(2\beta+1)} R_n(T_n, p) \geqslant C^*(1+o(1)), \qquad n \to +\infty, \tag{3.4}$$

where the infimum is taken over all estimators of p. This result can be found in Golubev (1992) without proof. The first proof of (3.4) should be attributed to Schipper (1996) who considered however, only integer values of β. Dalelane (2005) gives a proof of (3.4) which essentially follows that of Schipper (1996). In fact, it can be deduced from Golubev (1991) with additional arguments that are detailed in Rigollet (2004). Recently, Efromovich (2000) proved a general result for finding sharp minimax lower bounds in the multivariate case extending the one-dimensional setting of Efroimovich and Pinsker (1982). However, it cannot be applied in our set-up in full generality since it treats the densities with respect to a finite measure. From (3.3) and (3.4), we conclude that the Stein WGI estimator $\hat{p}_{\tilde{\lambda}}$ is sharp minimax adaptive over the scale of Sobolev classes of densities $\{\Theta(\beta, Q), \beta > 1/2, Q > 0\}$ in the sense of standard definitions given, for example, in Tsybakov (2004).

4. Application to kernel density estimation

For $h > 0$, define the kernel density estimator

$$\hat{p}_{n,h}(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{X_i - x}{h}\right),$$

where the kernel belongs to the class $\mathcal{K}_0$, the class of kernels K that admit a version of Fourier transform $\mathcal{F}[K]$ symmetric about 0, decreasing on $\mathbb{R}_+$ and taking its values in [0, 1]. Set $\Omega_n = n(\log n)^2$. Lemma 2.2, Corollary 2.4 and the fact that $\mathcal{F}[\hat{p}_{n,h}](\omega) = \varphi_n(\omega)\mathcal{F}[K](h\omega)$ together lead to the following theorem.

Theorem 4.1. *Assume that* $\int_{\mathbb{R}} |\varphi| \leqslant G$. *Then there exist an absolute constant* $C > 0$ *and a constant* $D_3 > 0$ *that depends only on G such that, for any* $\mu_n > C$ *and sufficiently large n, the Stein WGI estimator satisfies the kernel oracle inequality*

$$R_n(\hat{p}_{\tilde{\lambda}}, p) \leqslant \frac{1}{1 - C\mu_n^{-1}}\left((1 + \nu_n^{-1})(1 + \kappa_1(\log n)^{-1}) \inf_{K\in\mathcal{K}_0} \inf_{h>0} R_n(\hat{p}_{n,h}, p) + \frac{\tau_n(D_3)}{2\pi}\right). \quad (4.1)$$

Note that we do not suppose that $K \in L_1(\mathbb{R})$. This can be interpreted to mean that the infimum over $\mathcal{K}_0$ on the right-hand side of (4.1) is not attained for such kernels. The following lemma is similar to one in Stone (1984).

Lemma 4.1. *Let p be any density in* $L_2(\mathbb{R})$ *and* $K \in L_1(\mathbb{R})$ *be a symmetric kernel satisfying* $\int K(x)\mathrm{d}x = 1$ *and one of the following two conditions:*

(i) *K is non-negative;*
(ii) *K is a kernel of order 2s for a positive integer s (i.e. with moments* $\alpha_k \stackrel{\Delta}{=} \int t^k K(t)\mathrm{d}t = 0, \forall\, 1 \leqslant k < 2s$*) such that* $(-1)^{s+1}\alpha_{2s} > 0$ *and* $|K|$ *has a finite absolute moment of order* $2s + \delta$, $0 < \delta \leqslant 1$, *that is,* $\beta_{2s+\delta} \stackrel{\Delta}{=} \int |t|^{2s+\delta}|K(t)|\mathrm{d}t < \infty$.

Then there exists positive constants c, depending on p and K, and $a < 1$, *depending on K, such that*

$$\inf_{h>0} R_n(\hat{p}_{n,h}, p) \geqslant cn^{-a}. \quad (4.2)$$

Proof. The usual bias/variance decomposition is given by $R_n(\hat{p}_{n,h}, p) = \mathbb{E}_p\|\hat{p}_{n,h} - \mathbb{E}_p[\hat{p}_{n,h}]\|^2 + \|\mathbb{E}_p[\hat{p}_{h,h}] - p\|^2$. To bound the bias from below, we begin as in Stone's (1984) proof. According to the Plancherel identity,

$$2\pi\|\mathbb{E}_p[\hat{p}_{n,h}] - p\|^2 = \int_{\mathbb{R}} |\mathcal{F}[K](h\omega)\varphi(\omega) - \varphi(\omega)|^2 \,\mathrm{d}\omega = \int_{\mathbb{R}} (1 - \mathcal{F}[K](h\omega))^2 |\varphi(\omega)|^2 \,\mathrm{d}\omega.$$

Since φ is continuous and $\varphi(0) = 1$, there exists $\eta > 0$ such that $|\varphi(\omega)|^2 \geqslant 1/2$ for $|\omega| \leqslant \eta$. Then,

$$\int_{\mathbb{R}} (1 - \mathcal{F}[K](h\omega))^2 |\varphi(\omega)|^2 \,\mathrm{d}\omega \geqslant \frac{1}{2}\int_{-\eta}^{\eta} (1 - \mathcal{F}[K](h\omega))^2 \,\mathrm{d}\omega. \tag{4.3}$$

Suppose that K satisfies (i). Then it is a probability density and $\mathcal{F}[K]$ is its characteristic function. From Theorem 4.1.2 of Lukacs (1970), we obtain

$$\int_{-\eta}^{\eta} (1 - \mathcal{F}[K](h\omega))^2 \,\mathrm{d}\omega \geqslant \frac{1}{2^{4m}}\int_{-\eta}^{\eta} (1 - \mathcal{F}[K](2^m h\omega))^2 \,\mathrm{d}\omega, \qquad \forall m \in \mathbb{N}.$$

Now choose m such that $2^{-(m+1)} \leqslant h < 2^{-m}$:

$$\int_{-\eta}^{\eta} (1 - \mathcal{F}[K](h\omega))^2 \,\mathrm{d}\omega \geqslant h^4 \int_{-\eta/2}^{\eta/2} (1 - \mathcal{F}[K](\omega))^2 \,\mathrm{d}\omega = c_1 h^4,$$

where $c_1 = \int_{-\eta/2}^{\eta/2} (1 - \mathcal{F}[K](\omega))^2 \,\mathrm{d}\omega$ is a positive constant. Indeed if $c_1 = 0$ then, by continuity, $1 - \mathcal{F}[K](\omega) = 0$, for any $\omega \in (-\eta/2, \eta/2)$. Thus, by Theorem 4.1.1 of Lukacs (1970), $\mathcal{F}[K] \equiv 1$, which contradicts the conclusion of the Riemann–Lebesgue lemma. Therefore, there exists a positive constant c_1 such that

$$\|\mathbb{E}_p[\hat{p}_{n,h}] - p\|^2 \geqslant \frac{c_1}{4\pi}(h^4 \wedge 1), \qquad \forall h \in \mathbb{R}_+. \tag{4.4}$$

Suppose now that K satisfies (ii). By Theorem 2.2.1 of Lukacs (1983), since K is symmetric and of order $2s$, $\mathcal{F}[K](t)$ admits an expansion of the form

$$\mathcal{F}[K](t) = 1 + \sum_{k=1}^{s} (-1)^k \frac{\alpha_{2k} t^{2k}}{(2k)!} + O\big(|t|^{2s+\delta}\big) = 1 + (-1)^s \frac{\alpha_{2s} t^{2s}}{(2s)!} + O\big(|t|^{2s+\delta}\big),$$

as $|t| \to 0$. Note that Lukacs' theorem is stated for characteristic functions but it can be easily extended to our case. It is enough to suppose that $h \leqslant h_0$ for some positive constant h_0. If $h > h_0$, the bias is bounded from below by a positive constant and the lemma is trivially proved. For $h \leqslant h_0$ and $|\omega| \leqslant \eta$ with sufficiently small η, $1 - \mathcal{F}[K](h\omega) \geqslant (-1)^{s+1}\alpha_{2s} h^{2s}\omega^{2s}/[2(2s)!] > 0$. Therefore, for $h \leqslant h_0$,

$$\|\mathbb{E}_p[\hat{p}_{n,h}] - p\|^2 \geqslant c_2 h^{2s}, \tag{4.5}$$

where c_2 is a positive constant depending on K.

If $h \leqslant h_0$ for sufficiently small h_0, the variance term can be written (see Tsybakov 2004)

$$\mathbb{E}_p\|\hat{p}_{n,h} - \mathbb{E}_p[\hat{p}_{n,h}]\|^2 = \frac{1}{nh^2}\mathbb{E}_p\left[\int_{\mathbb{R}} K^2\left(\frac{X_1 - x}{h}\right)\mathrm{d}x - \left(\int_{\mathbb{R}} \mathbb{E}_p\left[K\left(\frac{X_1 - x}{h}\right)\right]\mathrm{d}x\right)^2\right]$$
$$\geqslant \frac{c_3}{nh}, \qquad c_3 > 0.$$

Using (4.4) and (4.5), we find that in both cases (i) and (ii) of the lemma there exists a positive constant $\tilde{a}$ such that

$$\inf_{h>0} R_n(\hat{p}_{n,h},\ p) \geqslant \inf_{h>0}\left[c_4 h^{\tilde{a}} + \frac{c_3}{nh}\right] \geqslant cn^{-a},$$

for positive constants c and $a < 1$. □

Corollary 4.1. *Let $K \in \mathcal{K}_0$ be a kernel satisfying the conditions of Lemma 4.1. Then for every fixed probability density $p \in L_2(\mathbb{R})$ with characteristic function $\varphi \in L_1(\mathbb{R})$,*

$$R_n(\hat{p}_{\tilde{\lambda}},\ p) \leqslant (1 + o(1))\inf_{h>0} R_n(\hat{p}_{n,h},\ p), \qquad n \to \infty. \tag{4.6}$$

Here $o(1)$ depends on K and p.

The proof is straightforward from Theorem 4.1 and Lemma 4.1.

It is important to note that the Stein WGI estimator mimics an oracle that is more powerful than any kernel oracle in a large class of kernels. The main difference from previous results of Stone's type (Stone 1984; Devroye and Penrod 1984; Wegkamp 1999) is that one and the same estimator is shown to be simultaneously as powerful as or even more powerful asymptotically than the kernel oracles corresponding to various kernels.

Some examples of admissible kernels covered by Theorem 4.1 and Corollary 4.1 are: the triangular kernel, the biweight kernel, Silverman's kernel, Fejer's kernel, the Gaussian kernel and the sinc kernel (the last one being covered only by Theorem 4.1 and not by Corollary 4.1). Note that in this list, which is not exhaustive, only the triangular and biweight kernels obey Stone's (1984) conditions. Of course, Theorem 4.1 says nothing about the kernels whose Fourier transform does not take values in [0, 1]. These kernels are not admissible (see Cline 1988) because they have higher MISE. This is the case of the Epanechnikov kernel and of the rectangular kernel.

5. Conclusion

In this paper, we propose a new density estimator called the Stein WGI estimator. The main result is an oracle inequality given in Theorem 2.4 showing that the estimator mimics the linear monotone oracle. We give two consequences of this inequality. First, we show that the Stein WGI estimator is sharp minimax adaptive over a scale of Sobolev classes of densities and second, that it is simultaneously as powerful as the kernel oracles corresponding to a large class of kernels.

6. Proof of Theorem 2.1

In all the lemmas presented in this section, we suppose that the assumptions of Theorem 2.1 are satisfied.

Let ζ_n be the empirical process

$$\zeta_n(\omega) \triangleq \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \big(\mathrm{e}^{\mathrm{i}\omega X_k} - \varphi(\omega)\big) = \sqrt{n}(\varphi_n(\omega) - \varphi(\omega)).$$

Several properties of this process have been given by Golubev and Levit (1996). We recall some of their results adapted to our framework. The following two lemmas are proved in Golubev and Levit (1996) for any φ such that $\int_{\mathbb{R}} |\varphi| \leqslant G$.

Lemma 6.1. *Let $n \geqslant 1$ be a fixed integer and b be an even real function such that*

$$\int_{\mathbb{R}} |b(\omega)| \mathrm{d}\omega \leqslant 2\sqrt{n}\|b\|, \tag{6.1}$$

where $\| . \|$ denotes the $L_2(\mathbb{R})$ norm. Then, for any $k \geqslant 1$, there exists a constant $C = C(G) > 0$ such that

$$\left| \mathbb{E}_p \int_{\mathbb{R}} b(\omega) \zeta_n(\omega) \mathrm{d}\omega \right|^{2k} \leqslant (Ck)^{4k} \|b\|^{2k}. \tag{6.2}$$

If, moreover, b satisfies

$$\max_{\substack{l \in \mathbb{N} \\ l \geqslant 3}} \|b\|^{-l} \int_{\mathbb{R}} |b(\omega)|^l \mathrm{d}\omega \leqslant 1, \tag{6.3}$$

then, for any integer $k \geqslant 1$, there exists a constant $C' = C'(G) > 0$ such that

$$\mathbb{E}_p \left(\int_{\mathbb{R}} b(\omega) \big(|\zeta_n(\omega)|^2 - \mathbb{E}_p |\zeta_n(\omega)|^2 \big) \mathrm{d}\omega \right)^{2k} \leqslant (C'k)^{4k} \|b\|^{2k}. \tag{6.4}$$

Lemma 6.2. *Let $\tilde{J} \in \{1, \ldots, N\}$ be a random index and $(\xi(\omega), \omega \in \mathbb{R})$ be a random process satisfying*

$$\mathbb{E}_p \left| \int_{\mathbb{R}} b_j(\omega) \xi(\omega) \mathrm{d}\omega \right|^{2k} \leqslant (Dk)^{4k} \|b_j\|^{2k}, \qquad k \in \mathbb{N}^*, \, j = 1, \ldots, N, \tag{6.5}$$

for some constant $D > 0$. Then there exists a constant $D' > 0$ such that

$$\mathbb{E}_p \left| \int_{\mathbb{R}} b_{\tilde{J}}(\omega) \xi(\omega) \mathrm{d}\omega \right| \leqslant (D' \log N)^2 \big(\mathbb{E}_p \|b_{\tilde{J}}\|^2 \big)^{1/2}.$$

The following lemma, stated by Golubev (1992), gives two important formulae that are used to construct oracle inequalities. The proof of this lemma can be found in Rigollet (2004).

Lemma 6.3. *Let A be a Borel subset of $\mathbb{R}$ and*

$$R_n^A(\hat{\varphi}_\lambda, \varphi) \triangleq \mathbb{E}_p \int_A |\hat{\varphi}_\lambda(\omega) - \varphi(\omega)|^2 \, \mathrm{d}\omega.$$

Then for all λ such that $\lambda(\omega) = \lambda(\omega, \mathbb{X}^n) = h(\mathbb{X}^n)\mathbb{1}_{\{\omega \in A\}} = h\mathbb{1}_{\{\omega \in A\}}$, $h \in \mathbb{R}$,

$$R_n^A(\hat{\varphi}_\lambda, \varphi) = \mathbb{E}_p[l_n(\lambda)] + \int_A |\varphi(\omega)|^2 \, \mathrm{d}\omega$$
$$+ \frac{2}{n}\mathbb{E}_p\left[\int_A h\big(|\zeta_n(\omega)|^2 - 1\big)\mathrm{d}\omega\right] + \frac{2}{\sqrt{n}}\mathrm{Re}\left[\mathbb{E}_p \int_A (h-1)\overline{\varphi(\omega)}\zeta_n(\omega)\mathrm{d}\omega\right] \tag{6.6}$$

and

$$\mathbb{E}_p[l_n(\lambda)] = \mathbb{E}_p\big[\Delta_n\big(\lambda, |\varphi|^2\big)\big] - \left(1 - \frac{1}{n}\right)\int_A |\varphi(\omega)|^2 \, \mathrm{d}\omega$$
$$+ \frac{1}{n}\mathbb{E}_p\left[\int_A (1-h)^2\big(|\zeta_n(\omega)|^2 - 1\big)\mathrm{d}\omega\right] + \frac{2}{\sqrt{n}}\mathrm{Re}\left[\mathbb{E}_p \int_A (1-h)^2\overline{\varphi(\omega)}\zeta_n(\omega)\mathrm{d}\omega\right]. \tag{6.7}$$

Lemma 6.4. *The Stein estimator on a Borel subset A of $\mathbb{R}$ satisfies the inequality*

$$R_n^A\Big(\hat{\varphi}_{\lambda_A^*}, \varphi\Big) \leqslant R_n^A\Big(\hat{\varphi}_{\lambda_A^{\mathrm{oracle}}}, \varphi\Big) + \frac{1}{n}\int_A |\varphi(\omega)|^2 \, \mathrm{d}\omega$$
$$+ \frac{2}{n}\mathbb{E}_p \int_A t_A^*\big(|\zeta_n(\omega)|^2 - \mathbb{E}_p|\zeta_n(\omega)|^2\big)\mathrm{d}\omega + \frac{2}{\sqrt{n}}\mathrm{Re}\left[\mathbb{E}_p \int_A (t_A^* - 1)\overline{\varphi(\omega)}\zeta_n(\omega)\mathrm{d}\omega\right]. \tag{6.8}$$

The proof of Lemma 6.4 is straightforward in view of the definitions of t_A^* and t_A^{oracle} and can be found in Rigollet (2004).

In view of Lemma 6.4, to prove Theorem 2.1, it is sufficient to bound from above the last two summands on the right-hand side of (6.8), that is,

$$\mathbb{E}_p \int_A t_A^*\big(|\zeta_n(\omega)|^2 - \mathbb{E}_p|\zeta_n(\omega)|^2\big)\mathrm{d}\omega \quad \text{and} \quad \mathbb{E}_p \int_A (t_A^* - 1)\overline{\varphi(\omega)}\zeta_n(\omega)\mathrm{d}\omega.$$

For this purpose, we will use Lemmas 6.1 and 6.2 with the additional assumption that A is symmetric about 0 as in Section 2.

Set $n \geqslant 1$ and let $|A| \geqslant 1$ be the length of A. Let us now cover the interval [0, 1] by $N = \lfloor n|A| \rfloor + 1 \leqslant 2n|A|$ disjoint intervals, $\Delta_1, \ldots, \Delta_N$, with centres $t_1, \ldots, t_N$ and lengths $1/(n|A|)$. Here $\lfloor x \rfloor$ denotes the integer part of x. Let $\tilde{t}_A$ be the projection of t_A^* on $\{t_1, \ldots, t_N\}$, that is,

$$\tilde{t}_A = \arg \min_{t \in \{t_1, \dots, t_N\}} |t - t_A^*|.$$

Thus $|\tilde{t}_A - t_A^*| \leqslant 1/n|A|$ and we have

$$\left| \mathbb{E}_p \int_A (t_A^* - 1)\overline{\varphi(\omega)}\zeta_n(\omega)\mathrm{d}\omega \right| \leqslant \left| \mathbb{E}_p \int_A (\tilde{t}_A - 1)\overline{\varphi(\omega)}\zeta_n(\omega)\mathrm{d}\omega \right| + \left| \mathbb{E}_p \int_A (t_A^* - \tilde{t}_A)\overline{\varphi(\omega)}\zeta_n(\omega)\mathrm{d}\omega \right|.$$

Next, we have

$$\begin{aligned} \left| \mathbb{E}_p \int_A (t_A^* - \tilde{t}_A)\overline{\varphi(\omega)}\zeta_n(\omega)\mathrm{d}\omega \right| &\leqslant \frac{1}{n|A|} \int_A \mathbb{E}_p |\varphi(\omega)|\,|\zeta_n(\omega)|\mathrm{d}\omega \\ &\leqslant \frac{1}{2n|A|} \int_A \mathbb{E}_p \big(|\varphi(\omega)|^2 + |\zeta_n(\omega)|^2\big)\mathrm{d}\omega = \frac{1}{2n}. \end{aligned}$$

By the same argument,

$$\left| \mathbb{E}_p \int_A (t_A^* - \tilde{t}_A)\big(|\zeta_n(\omega)|^2 - \mathbb{E}_p|\zeta_n(\omega)|^2\big)\mathrm{d}\omega \right| \leqslant \frac{2}{n}.$$

Note now that $b_j^{\mathrm{Re}}(\omega) \stackrel{\Delta}{=} (t_j - 1)\mathrm{Re}\big(\overline{\varphi(\omega)}\big)\mathbb{1}_{\{\omega \in A\}}$ satisfies (6.1), for $|A| \leqslant 4n$ and $j = 1, \dots, N$. Indeed, by the Cauchy–Schwarz inequality,

$$\begin{aligned} \int_{\mathbb{R}} |b_j^{\mathrm{Re}}(\omega)|\mathrm{d}\omega &= \int_{\mathbb{R}} |(t_j - 1)\mathrm{Re}(\varphi(\omega))\mathbb{1}_{\{\omega \in A\}}|\mathrm{d}\omega \\ &\leqslant (1 - t_j)\sqrt{|A|}\left(\int_A |\mathrm{Re}(\varphi(\omega))|^2\,\mathrm{d}\omega \right)^{1/2} \\ &\leqslant 2\sqrt{n}(1 - t_j)\left(\int_A |\mathrm{Re}(\varphi(\omega))|^2\,\mathrm{d}\omega \right)^{1/2} \end{aligned}$$

and

$$\|b_j^{\mathrm{Re}}\| = \left(\int_{\mathbb{R}} |(1 - t_j)\mathrm{Re}(\varphi(\omega))\mathbb{1}_{\{\omega \in A\}}|^2\,\mathrm{d}\omega \right)^{1/2} = (1 - t_j)\left(\int_A |\mathrm{Re}(\varphi(\omega))|^2\,\mathrm{d}\omega \right)^{1/2},$$

hence

$$\int_{\mathbb{R}} |b_j^{\mathrm{Re}}(\omega)|\mathrm{d}\omega \leqslant 2\sqrt{n}\|b_j^{\mathrm{Re}}\|.$$

In the same manner, $b_j^{\mathrm{Im}}(\omega) \stackrel{\Delta}{=} (t_j - 1)\mathrm{Im}\big(\overline{\varphi(\omega)}\big)\mathbb{1}_{\{\omega \in A\}}$ satisfies (6.1).

Moreover, $b_j'(\omega) \stackrel{\Delta}{=} t_j\mathbb{1}_{\{\omega \in A\}}$ satisfies (6.1) and (6.3) for $1 \leqslant |A| \leqslant 4n$ and $j = 1, \dots, N$. Indeed,

$$\int_{\mathbb{R}} |b_j'(\omega)|\mathrm{d}\omega = \int_{\mathbb{R}} |t_j\mathbb{1}_{\{\omega \in A\}}|\mathrm{d}\omega = t_j|A| \leqslant 2\sqrt{n}t_j\sqrt{|A|}$$

and

$$\|b_j'\| = \left(\int_{\mathbb{R}} |t_j \mathbb{1}_{\{\omega\in A\}}|^2 \,\mathrm{d}\omega\right)^{1/2} = t_j\sqrt{|A|},$$

hence

$$\int_{\mathbb{R}} |b_j'(\omega)|\mathrm{d}\omega \leqslant 2\sqrt{n}\|b_j'\|.$$

Therefore $b_j'(\omega) = t_j \mathbb{1}_{\{\omega\in A\}}$ satisfies (6.1). On the other hand,

$$\|b_j'\| = t_j\sqrt{|A|} \quad \text{and} \quad \int_{\mathbb{R}} |b_j'(\omega)|^l \mathrm{d}\omega = (t_j)^l |A|.$$

Therefore, for all $l \geqslant 2$, and for $|A| \geqslant 1$,

$$\|b_j'\|^{-l}\int_{\mathbb{R}} |b_j'(\omega)|^l \mathrm{d}\omega = |A|^{1-l/2} \leqslant 1.$$

Thus, $b_j'(\omega) = t_j \mathbb{1}_{\{\omega\in A\}}$ satisfies (6.3). By Lemma 6.1 for any integer $k \geqslant 1$, any $j = 1, \ldots, N$ and any A such that $1 \leqslant |A| \leqslant 4n$, we have the following upper bounds:

$$\mathbb{E}_p\left|\int_A (\tilde{t}_A - 1)\overline{\varphi(\omega)}\zeta_n(\omega)\mathrm{d}\omega\right|^{2k} \leqslant (Ck)^{4k}\left((1-t_j)^2\int_A |\varphi(\omega)|^2\,\mathrm{d}\omega\right)^k \tag{6.9}$$

and

$$\mathbb{E}_p\left(\int_A t_j\big(|\zeta_n(\omega)|^2 - \mathbb{E}_p|\zeta_n(\omega)|^2\big)\mathrm{d}\omega\right)^{2k} \leqslant (C'k)^{4k}\Big(t_j\sqrt{|A|}\Big)^{2k}. \tag{6.10}$$

By (6.9) and (6.10), Lemma 6.2 can be applied to processes $\xi(\omega) = \zeta_n(\omega)$ and $\xi(\omega) = |\zeta_n(\omega)|^2 - \mathbb{E}_p|\zeta_n(\omega)|^2$. We obtain the inequalities

$$\mathbb{E}_p\left|\int_A (\tilde{t}_A - 1)\overline{\varphi(\omega)}\zeta_n(\omega)\mathrm{d}\omega\right| \leqslant C(\log N)^2\left(\mathbb{E}_p\big[(1-\tilde{t}_A)^2\big]\int_A |\varphi(\omega)|^2\,\mathrm{d}\omega\right)^{1/2} \tag{6.11}$$

and

$$\mathbb{E}_p\left|\int_A \tilde{t}_A\big(|\zeta_n(\omega)|^2 - \mathbb{E}_p|\zeta_n(\omega)|^2\big)\mathrm{d}\omega\right| \leqslant C'(\log N)^2\big(\mathbb{E}_p\big[\tilde{t}_A^{\,2}\big]|A|\big)^{1/2}. \tag{6.12}$$

Furthermore, for $|A| \geqslant 1$ and any $\mu_n > 0$,

$$\begin{aligned}\frac{2}{\sqrt{n}}\left(\mathbb{E}_p\big[(1-\tilde{t}_A)^2\big]\int_A |\varphi(\omega)|^2\,\mathrm{d}\omega\right)^{1/2} &\leqslant \frac{(\log n)^2\mu_n}{n} + \frac{1}{(\log n)^2\mu_n}\mathbb{E}_p\big[(1-\tilde{t}_A)^2\big]\int_A |\varphi(\omega)|^2\,\mathrm{d}\omega \\ &\leqslant C\frac{(\log n)^2\mu_n}{n} + \frac{1}{(\log n)^2\mu_n}\mathbb{E}_p\big[(1-t_A^*)^2\big]\int_A |\varphi(\omega)|^2\,\mathrm{d}\omega.\end{aligned}$$

And, by the same argument

$$\frac{2}{n}\left(\mathbb{E}_p\left[\tilde{t}_A^{\,2}\right]|A|\right)^{1/2} \leqslant C'\frac{(\log n)^2\mu_n}{n} + \frac{1}{n(\log n)^2\mu_n}\mathbb{E}_p\left[(t_A^*)^2\right]\int_A\left(1-|\varphi(\omega)|^2\right)\mathrm{d}\omega$$
$$+\frac{1}{n(\log n)^2\mu_n}\int_A|\varphi(\omega)|^2\,\mathrm{d}\omega.$$

Therefore, using the two preceding inequalities, Lemma 6.4, (6.11), (6.12) and the residuals of order $1/n$ due to discretization, one obtains, for strictly positive constants C_1, C_2 and C_3,

$$R_n^A\left(\hat{\varphi}_{\lambda_A^*},\varphi\right) \leqslant R_n^A\left(\hat{\varphi}_{\lambda_A^{\mathrm{oracle}}},\varphi\right) + \frac{1}{n}\int_A|\varphi(\omega)|^2\,\mathrm{d}\omega + \frac{1}{n(\log n)^2\mu_n}\int_A|\varphi(\omega)|^2\,\mathrm{d}\omega$$
$$+ C_1\frac{(\log N)^2}{(\log n)^2\mu_n}R_n^A\left(\hat{\varphi}_{\lambda_A^*},\varphi\right) + C_2\frac{(\log N)^2(\log n)^2\mu_n}{n} + \frac{C_3}{n}.$$

Then, noting that

$$\log N \leqslant \log(2|A|n) \leqslant \log(8n^2) \leqslant 5\log n, \qquad \text{for } n \geqslant 2,$$

we find a positive constant C_4 such that,

$$\left(1 - C_3\frac{1}{\mu_n}\right)R_n^A\left(\hat{\varphi}_{\lambda_A^*},\varphi\right) \leqslant R_n^A\left(\hat{\varphi}_{\lambda_A^{\mathrm{oracle}}},\varphi\right) + \frac{2}{n}\int_A|\varphi(\omega)|^2\,\mathrm{d}\omega + C_4\frac{(\log n)^4\mu_n}{n}. \tag{6.13}$$

and Theorem 2.1 is proved.

Acknowledgement

I am grateful to Alexander Tsybakov for introducing me to the domain and for his constant help and support. I would also like to thank Cristina Butucea, Arnak Dalalyan and Yuri Golubev for comments that helped to improve the text.

References

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

Boiko, L.L. and Golubev, G.K. (2000) How to improve the nonparametric density estimation in S-PLUS. *Problems Inform. Transmission*, **36**, 354–361.

Cai, T. (1999) Adaptive wavelet estimation: a block thresholding and oracle inequality approach. *Ann. Statist.*, **27**, 898–924.

Cavalier, L. and Tsybakov, A.B. (2001) Penalized blockwise Stein's method, monotone oracles and sharp adaptive estimation. *Math. Methods Statist.*, **10**, 247–282.

Cavalier, L., Golubev, Yu., Picard, D. and Tsybakov, A.B. (2002) Oracle inequalities for inverse problems. *Ann. Statist.*, **30**, 843–874.

Cline, D.B.H. (1988) Admissible kernel estimators of a multivariate density. *Ann. Statist.*, **16**, 1421–1427.

Dalelane, C. (2005) Exact minimax risk for density estimators in non-integer Sobolev classes. Preprint no. 979, Laboratoire de Probabilités et Modèles Aléatoires, Universités Paris 6 and Paris 7 (http://www.proba.jussieu.fr/mathdoc/preprints).

Devroye, L. and Penrod, C.S. (1984) Distribution-free lower bounds in density estimation. *Ann. Statist.*, **12**, 1250–1262.

Donoho, D.L. and Johnstone, I.M. (1994) Ideal spatial adaptation via wavelet shrinkage. *Biometrika*, **81**, 425–455.

Donoho, D.L. and Johnstone, I.M. (1995) Adapting to unknown smoothness via wavelet shrinkage. *J. Amer. Statist. Assoc.*, **90**, 1200–1224.

Efroimovich, S.Yu. (1985) Nonparametric estimation of a density with unknown smoothness. *Theory Probab. Appl.*, **30**, 557–568.

Efroimovich, S.Yu. and Pinsker, M.S. (1982) Estimation of square-integrable probability density of a random variable *Problems Inform. Transmission*, **18**, 175–189.

Efroimovich, S.Yu. and Pinsker, M.S. (1984) Learning algorithm for nonparametric filtering. *Automat. Remote Control*, **11**, 1434–1440.

Efromovich, S.Yu. (1999) *Nonparametric Curve Estimation.* New York: Springer-Verlag.

Efromovich, S. (2000) On sharp adaptive estimation of multivariate curves. *Math. Methods Statist.*, **9**, 117–139.

Efromovich, S. (2004a) Oracle inequalities for Efromovich–Pinsker blockwise estimates. *Methodol. Comput. Appl. Probab.*, **6**, 303–322.

Efromovich, S. (2004b) Analysis of blockwise shrinkage wavelet estimates via lower bounds for no-signal setting. *Ann. Inst. Statist. Math.*, **56**, 205–223.

Efromovich, S. (2005) Estimation of the density of regression errors. *Ann. Statist.*, **33**, 2194–2227.

Ekeland, I. and Temam, R. (1974) *Analyse convexe et problèmes variationnels.* Paris: Dunod Gauthier-Villars.

Goldenshluger, A., and Tsybakov, A. (2001) Adaptive prediction and estimation in linear regression with infinitely many parameters. *Ann. Statist.*, **29**, 1601–1619.

Golubev, G.K. (1990) Quasi-linear estimates of signals in L_2. *Problems Inform. Transmission*, **26**, 15–20.

Golubev, G.K. (1991) LAN in nonparametric estimation of functions and lower bounds for quadratic risks. *Theory Probab. Appl.*, **36**, 152–157.

Golubev, G.K. (1992) Nonparametric estimation of smooth probability densties in L_2. *Problems Inform. Transmission*, **28**, 44–54.

Golubev, G.K. and Levit B.Y. (1996) Distribution function estimation: adaptive smoothing. *Math. Methods Statist.*, **5**, 383–403.

Golubev, G.K. and Nussbaum, M. (1992) Adaptive spline estimates in a nonparametric regression model. *Theory Probab. Appl.*, **37**, 521–529.

Hall, P., Kerkyacharian, G. and Picard, D. (1998) Block thresholding rules for curve estimation using kernel and wavelet methods. *Ann. Statist.*, **26**, 922–942.

Härdle, W., Kerkyacharian, G., Picard, D., and Tsybakov, A. (1998) *Wavelets, Approximation, and Statistical Applications*, Lecture Notes in Statist. 129. New York: Springer-Verlag.

Kneip, A. (1994) Ordered linear smoothers. *Ann. Statist.*, **22**, 835–866.

Li, K.-C. (1987) Asymptotic optimality of C_P, C_L, cross-validation and generalized cross-validation: Discrete index set. *Ann. Statist.*, **15**, 958–976.

Lukacs, E. (1970). *Characteristic Functions*, 2nd edn. London: Griffin.

Lukacs, E. (1983) *Developments in Characteristic Function Theory.* London: Griffin.

Nemirovski, A. (2000) Topics in non-parametric statistics. In P. Bernard (ed.), *Lectures on Probability*

Theory and Statistics. École d'Été de Probabilités de Saint-Flour XXVIII – 1998, Lecture Notes in Math. 1738, pp. 85–277. Berlin: Springer-Verlag.

Polyak, B.T. and Tsybakov, A.B. (1992) A family of asymptotic optimal methods for choosing the estimate order in orthogonal series regression. *Theory Probab. Appl.*, **37**, 471–481.

Rigollet, R. (2004) Adaptive density estimation using Stein's blockwise method. Preprint no. 913, Laboratoire de Probabilités et Modèles Aléatoires, Universités Paris 6 and Paris 7 (http://www.proba.jussieu.fr/mathdoc/preprints).

Schipper, M. (1996) Optimal rates and constants in L_2-minimax estimation of probability density functions. *Math. Methods Statist.*, **5**, 253–274.

Shibata, R. (1981) An optimal selection of regression variables. *Biometrika*, **68**, 45–54.

Stone, C.J. (1984) An asymptotically optimal window selection rule for kernel density estimates. *Ann. Statist.*, **12**, 1285–1297.

Tsybakov, A. (2002) Discussion of 'Random rates in anisotropic regression' by M. Hoffmann and O. Lepski. *Ann. Statist.*, **30**, 379–385.

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

Wegkamp, M.H. (1999) Quasi-universal bandwidth selection for kernel density estimators. *Canad. J. Statist.*, **27**, 409–420.

Received April 2004 and revised October 2005