On Spatial Adaptive Estimation of Nonparametric Regression

A. Goldenshluger and A. Nemirovski

Faculty of Industrial Engineering & Managament, Technion,
Israel Institute of Technology, Haifa, Israel.
e–mail: goldensh@ie.technion.ac.il, nemirovs@ie.technion.ac.il

Abstract

The paper is devoted to developing spatial adaptive estimates for restoring functions from noisy observations. We show that the traditional least square (piecewise polynomial) estimate equipped with adaptively adjusted window possesses simultaneously many attractive adaptive properties, namely, 1) it is near–optimal within $\ln n$–factor for estimating a function (or its derivative) at a single point; 2) it is *spatial adaptive* in the sense that its quality is close to that one which could be achieved if smoothness of the underlying function was known in advance; 3) it is optimal in order (in the case of "strong" accuracy measure) or near–optimal within $\ln n$–factor (in the case of "weak" accuracy measure) for estimating whole function (or its derivative) over wide range of the classes and global loss functions. We demonstrate that the "spatial adaptive abilities" of our estimate are, in a sense, the best possible. Besides this, our adaptive estimate is computationally efficient and demonstrates reasonable practical behavior.

1 Introduction

Suppose we are given noisy observations $y(x)$ of a *signal* – a function $f : [0,1] \to \mathbf{R}$ – along the regular grid $\Gamma_n = \{i/n, i = 0, \dots, n\}$:

$$y(x) = f(x) + \xi(x), \qquad x \in \Gamma_n, \tag{1}$$

where $\{\xi(x)\}_{x\in\Gamma_n}$ is a sequence of independent $\mathcal{N}(0,1)$ random variables defined on the underlying probability space $(\Omega, \mathcal{A}, P)$. The problem we are interested in is to restore f. An *estimate*, by definition, is a Borel in $x \in [0,1]$ and in the observations $y^n = \{y(x)\}_{x\in\Gamma_n}$ function $\hat{f}(x, y^n)$. Given an estimate, we can measure its quality *at a fixed* f by various ways. If we are interested in the value of f at a given point x_0, then it is natural to measure the quality by *local risks* of the type

$$\mathcal{R}_{x_0;r}(\hat{f}, f) = \left[\mathcal{E}|\hat{f}(x_0) - f(x_0)|^r\right]^{1/r}, \ 1 \le r < \infty$$

(from now on the expectation $\mathcal{E}$ is taken over the distribution of $\xi(\cdot)$). We can be also interested in the quality of restoring f at a given segment $\Delta \subset [0,1]$, and in this case it is natural to use *global risks* of the type

$$\mathcal{R}_{q,\Delta;r}(\hat{f}, f) = \left[\mathcal{E}\|\hat{f} - f\|^r_{q,\Delta}\right]^{1/r}, 1 \le r < \infty,$$

where $1 \leq q \leq \infty$ and $\| \cdot \|_{q,\Delta}$ is the usual L_q-norm on Δ.

The above measures of quality of an estimate depend not only on the estimate itself, but also on f. The standard way to eliminate the dependence on f and thus to come to a situation where one can think about the quality of the estimate itself is to pass from risks at a fixed f to the uniform risks over a given family $\mathcal{F}$ of signals, i.e., to the quantities

$$\mathcal{R}_{x_0;r}(\hat{f}, \mathcal{F}) = \sup_{f \in \mathcal{F}} \mathcal{R}_{x_0;r}(\hat{f}, f)$$

or

$$\mathcal{R}_{q,\Delta;r}(\hat{f}, \mathcal{F}) = \sup_{f \in \mathcal{F}} \mathcal{R}_{q,\Delta;r}(\hat{f}, f).$$

For a fixed family $\mathcal{F}$ and a specified accuracy measure (i.e., $\mu = (x_0; r)$ or $\mu = (q, \Delta; r)$) one can define the *optimal* minimax risks

$$\mathcal{R}^*_\mu(n; \mathcal{F}) = \inf_{\hat{f}} \mathcal{R}_\mu(\hat{f}, \mathcal{F}),$$

and pose the problem of finding an *optimal in order* method of estimation, i.e., a sequence $\hat{f}_n$ of estimates with

$$\mathcal{R}_\mu(\hat{f}_n, \mathcal{F}) \leq C(n) \mathcal{R}^*_\mu(n; \mathcal{F}),$$

$\sup_n C(n) < \infty$.

The outlined *nonparametric regression problem* is the subject of considerable literature, where advanced results were obtained for variety of classes $\mathcal{F}$ (usually defined by specifying smoothness of signals) and of the accuracy measures μ. Note, anyhow, that the majority of estimates known from the literature heavily depend on the class $\mathcal{F}$ the estimate is designed for (and sometimes also on the accuracy measure in question). In practice, however, we normally do not know in advance what are the classes $\mathcal{F}$ the actual signal belongs to, and thus meet with severe difficulties when trying to use "highly specialized" estimates (for detailed discussion of these conceptual issues, see Donoho *et. al.* [3]). As a result, there is a strong interest in developing *adaptive estimates* which are optimal in order not over a single class $\mathcal{F}$, but over a family of these classes (the wider is the family, the more attractive, from the practical viewpoint, is the estimate).

The breakthrough in constructing adaptive estimates is due to Pinsker, Efroimovich [9] who developed an optimal in order estimator (even with $C(n) = 1 + o(1)$ as $n \to \infty$), with respect to the accuracy measure $\mu = (2, [0, 1], 2)$, over all ellipsoids

$$\mathcal{F} = \{ f(\cdot) = \sum_{i=1}^{\infty} f_i e_i(\cdot) : \sum_{i=1}^{\infty} a_i^2 f_i^2 \leq R^2 \}$$

associated with nondecreasing sequences $a_i \to \infty$ and the standard trigonometric Fourier basis $\{e_i\}$ in $L_2[0, 1]$. Thus, when interested in L_2 losses, one can adapt to the unknown smoothness (also measured in the L_2 scale), at least in the case of smooth periodic signals (the periodicity assumption was later eliminated). In contrast to this, for the problem of estimating at a given point, Lepskii [5], along with developing kernel estimator with adaptively adjusted smoothing parameter, has shown that it is impossible to adapt a "point" estimator to the (unknown) smoothness of the signal, and here only "$\ln n$-spoiled", as compared to the case of known smoothness, risks can be achieved; for general theory of adaptive nonparametric estimation, see Lepskii [6].

From practical viewpoint, an extremely important property of a "good" estimator is its *spatial adaptivity*, the property as follows: if there exists a segment $\bar{\Delta} \subset [0,1]$ which covers the point x_0 (or a segment Δ) where we are estimating f, and f is smooth with certain parameters of smoothness on $\bar{\Delta}$, then the quality of the estimator should be at least close to that one which could be achieved if we knew in advance $\bar{\Delta}$ and the related parameters of smoothness of f on this segment.

Recently, several spatial adaptive estimates were proposed (the wavelet-based estimators of Donoho et al. [2, 3], and Juditsky [4], adaptive kernel estimates of Lepskii, Mammen and Spokoiny [7]); in the cited papers, the smoothness of the signal is specified as membership in the Besov or Triebel spaces. The goal of this paper is to develop an estimator which is spatial adaptive with respect to the usual Sobolev spaces. The estimator is extremely simple and does not use the wavelet technique; in fact it is a particular implementation of the general scheme of Lepskii. The local and global risks of our estimator, over a wide range of characterizations of smoothness and accuracy measures, are only by logarithmic in n factor worse than the corresponding optimal risks; this logarithmic factor disappears for global risks related to large values of q, and for the case of "small" values of q is, in a sense, an unavoidable price for spatial adaptation. Along with these theoretical properties, the estimator in question demonstrates quite reasonable practical behavior.

The main body of the paper is organized as follows. Section 2 explains the idea of the estimator; the estimator itself is constituted in Section 3, where we also investigate its local risks. The global risks are studied in Section 4. Section 5 extends the construction to the case when we are interested in estimating not only the values, but also the derivatives of the signal. Section 7 presents the numerical results obtained on the test problems from [1, 2]. Section 6 demonstrates that the spatial adaptive abilities of the developed estimate are, in a sense, the best possible, even in the case $p = q$. Section 8 contains some concluding remarks.

2 The idea

The idea of our construction is as follows. Assume that we are interested in estimating f at a given point $x_0 \in (0,1)$ and that all we know in advance is that f is continuous at x_0. Then the natural way to estimate $f(x_0)$ is to use the simplest *window average*

$$\hat{f}_\Delta(x_0) = \frac{1}{N_\Delta} \sum_{x \in M_\Delta} y(x),$$

where *the window* $\Delta \in [0,1]$ is some segment $[x_0 - \delta, x_0 + \delta]$ centered at x_0 and containing at least one observation point, M_Δ is the set of observation points in Δ and N_Δ is the cardinality of M_Δ. The problem, of course, is how to choose "the best" window when no a priori information on f is available. To answer this question, note that the natural upper bound for the error $|\hat{f}_\Delta(x_0) - f(x_0)|$ is as follows:

$$|\hat{f}_\Delta(x_0) - f(x_0)| \le \omega_f(x_0,\delta) + N_\Delta^{-1/2}|\zeta(\Delta)|, \;\; \zeta(\Delta) = \frac{1}{\sqrt{N_\Delta}} \sum_{x \in M_\Delta} \xi(x), \tag{2}$$

where $\omega_f(x,\delta) = \sup_{x \in \Delta} |f(x) - f(x_0)|$. The right hand side of (2) is comprised of two terms - the deterministic *dynamic error* $\omega_f(x_0,\delta)$ which, for a given δ, is completely determined by f, and the *stochastic error* $N_\Delta^{-1/2}|\zeta(\Delta)|$ which is completely independent of f. Since $\zeta(\Delta)$ is $\mathcal{N}(0,1)$, the stochastic error typically is of order of $(n\delta)^{-1/2}$:

$$P\{N_\Delta^{-1/2}|\zeta(\Delta)| > \kappa(n\delta)^{-1/2}\} \le \exp\{-c\kappa^2\} \tag{3}$$

with certain absolute constant $c > 0$. Now, there are no more than n essentially different (resulting in different sets M_Δ) choices of Δ; let these choices be $\Delta_1 \subset \Delta_2 \subset \ldots \subset \Delta_N$, and let $2\delta_1 \le \ldots \le 2\delta_N$ be the lengths of the windows Δ_1,..., Δ_N. Let

$$\Omega_\kappa = \{\omega \in \Omega \mid N_{\Delta_i}^{-1/2}|\zeta(\Delta_i)| \le \kappa(n\delta_i)^{-1/2},\ i = 1, ..., N\};$$

then, by (3),

$$P(\Omega\backslash\Omega_\kappa) \le N\exp\{-c\kappa^2\} \le n\exp\{-c\kappa^2\}. \tag{4}$$

Assume that $\omega \in \Omega_\kappa$. Then (2) can be strengthen as

$$|\hat{f}_i(x_0) - f(x_0)| \le \omega_f(x_0, \delta_i) + \kappa(n\delta_i)^{-1/2},\ \hat{f}_i \equiv \hat{f}_{\Delta_i}. \tag{5}$$

As i (and, consequently, δ_i) grows, the first term in the right hand side increases, and the second term decreases; therefore a reasonable choice of the window to be used is that one which balances both the terms, say, the one related to the largest i with $\omega_f(x_0, \delta_i) \le \kappa(n\delta_i)^{-1/2}$. Let us denote by $i^* = i^*_\kappa(f)$ the indicated value of i, and let us call the quantity

$$\rho^*_f(n) = (n\delta_{i^*})^{-1/2}$$

the *ideal risk*.

It is easily seen that the ideal risk in fact is ideal – for any class of the type

$$\mathcal{F} = \{f \mid \left(\int_{x_0-d}^{x_0+d} |f'(x)|^p dx\right)^{1/p} \le L\}$$

associated with fixed d, p, L, $\sup_{f\in\mathcal{F}} \rho^*_f(n)$ is at most of the order of the optimal for the class $\mathcal{F}$ local risk as $n \to \infty$. The crucial point is that *there exists an estimate* $\hat{f}(x_0)$ *with the inaccuracy, for the case of* $\omega \in \Omega_\kappa$, *of order of* κ *times the ideal risk.* To get $\hat{f}$, let us act as follows. Let $\omega \in \Omega_\kappa$. When $i \le i^*_\kappa(f)$, the inaccuracy of the estimate $\hat{f}_i$ satisfies (5) and is therefore at most $2\rho_i$,

$$\rho_i = \kappa(n\delta_i)^{-1/2},$$

since the "dynamic" term $\omega_f(x_0, \delta_i)$ for these i is dominated by the "stochastic" term $\kappa(n\delta_i)^{-1/2}$. It follows that

(*): all the segments

$$D_i = [\hat{f}_i(x_0) - 2\rho_i, \hat{f}_i(x_0) + 2\rho_i]$$

corresponding to $i \le i^*_\kappa(f)$ have a point in common, namely, $f(x_0)$.

Now let i^+ be the largest of those i for which the segments D_j, $j \le i$, have a point in common; i^+ *is defined in terms of* $\hat{f}_i$, $i = 1, 2, ..., N$, *and* κ *only, so that*

$$\hat{f}(x_0) \equiv \hat{f}_{i^+}(x_0)$$

is an estimate – it is defined in terms of observations (and κ*).* We claim that for $\omega \in \Omega_\kappa$ inaccuracy of the estimate $\hat{f}(x_0)$ is at most 6κ times the ideal risk. Indeed, if $\omega \in \Omega_\kappa$ then, as we know from (*), the segments D_i, $i \le i^* \equiv i^*_\kappa(f)$, have a point in common; whence $i^+ \ge i^*_\kappa(f)$ and D_{i^+} intersects D_{i^*} (by definition of i^+). By the same (*), D_{i^*} covers $f(x_0)$. Combining these observations, we see that the distance between $\hat{f}$ (i.e., the midpoint of the segment D_{i^+}) and $f(x_0)$ is at most the length $4\rho_{i^*}$ of the segment D_{i^*} plus the half-length $2\rho_{i^+}$ of the segment D_{i^+}, so that

$$|\hat{f}(x_0) - f(x_0)| \le 4\rho_{i^*} + 2\rho_{i^+} \le 6\rho_{i^*} \equiv 6\kappa\rho^*_f(n)$$

(note that $i^+ \ge i^*$ and therefore $\rho_{i^+} \le \rho_{i^*}$), as claimed.

Thus, in the case of $\omega \in \Omega_\kappa$ quality of the estimate $\hat{f}(x_0)$ is of order of κ times the ideal risk, and the latter, as it was already mentioned, is of order of the optimal risk on any class of functions given by first order smoothness restrictions on f. Of course, when $\omega \notin \Omega_\kappa$, then the local risk of the aforementioned estimate can be much larger than the ideal risk; recall, anyhow, that probability of the "bad" event $\omega \notin \Omega_\kappa$ is of order of $\exp\{-O(\kappa^2)\}$ (see (4)); choosing $\kappa = O(\sqrt{\ln n})$, we make the influence of this bad event negligibly small, and cost of this "suppressing fluctuations" is the appearance of a logarithmic in n factor at the ideal risk.

3 The estimate

3.1 Construction

Let us consider the following class of estimates of f at a given point $u \in [0,1]$. We fix a nonnegative integer m – *the order* of the estimate – and fit a polynomial of the degree m to the observations in some *window* (a segment contained in $[0,1]$ and containing at least $m+1$ observation points) Δ by the least square method. As an estimate of $f(u)$, we take the value of the polynomial at the point u. Such the estimate is designated by $\mathcal{S}^m_{\Delta,u}(y^n)$ and is referred to as *Least Square Estimate* of the order m associated with Δ and u. Formally, let $\mathcal{W}$ be the set of all segments from $[0,1]$ containing at least $m+1$ observation points. For any $\Delta \in \mathcal{W}$ and any $u \in [0,1]$ the estimate $\mathcal{S}^m_{\Delta,u}(y^n)$ is defined as

$$\mathcal{S}^m_{\Delta,u}(y^n) = \sum_{x \in M_\Delta} \alpha_\Delta(x,u) y(x), \tag{6}$$

where $M_\Delta \equiv \Gamma_n \cap \Delta$, and the coefficients $\{\alpha_\Delta(x,u)\}_{x\in M_\Delta}$ form the solution to the following optimization problem

$$\begin{array}{rrcl} & \sum_{x\in M_\Delta} \alpha^2_\Delta(x,u) & = & min \\ subject\ to & \sum_{x\in M_\Delta} \alpha_\Delta(x,u) x^j & = & u^j, \quad j = 0,\dots,m. \end{array} \tag{7}$$

Let us call the quantity

$$r_{\Delta,u} = \left(\sum_{x\in M_\Delta} \alpha^2_\Delta(x,u) \right)^{1/2}$$

the index of the window Δ at the point u. It is well-known that for any window $\Delta \in \mathcal{W}$ and any $u \in \Delta$ one has

$$|\alpha_\Delta(x,u)| \le \frac{\theta_m}{N_\Delta},\ x \in M_\Delta, \tag{8}$$

N_Δ being the cardinality of M_Δ, θ_m being a constant depending on m only, and, consequently,

$$r_{\Delta,u} \le \frac{\theta_m}{\sqrt{N_\Delta}}. \tag{9}$$

To specify the adaptive estimate we need the following definitions.

Definition 1 *Let $x_0 \in (0,1)$. A window $\Delta \in \mathcal{W}$ is called admissible for x_0, if it is centered at x_0.*

The set of all windows admissible for x_0 will be denoted $\mathcal{D}(x_0)$.

Definition 2 *Let $\kappa \ge 1$. A window $\Delta \in \mathcal{D}(x_0)$ is called κ-good, if the set*

$$\mathcal{I}_\Delta \equiv \bigcap_{\substack{\Delta' \in \mathcal{D}(x_0) \\ \Delta' \subseteq \Delta}} \left[\mathcal{S}^m_{\Delta',x_0} - 2\kappa r_{\Delta',x_0},\, \mathcal{S}^m_{\Delta',x_0} + 2\kappa r_{\Delta',x_0}\right]$$

is nonempty.

Note that κ-good windows do exist (e.g., the minimal with respect to inclusion window $\Delta_{min}(x_0)$ in the family $\mathcal{D}(x_0)$ clearly is κ-good). It is also clear that there exists the maximal with respect to inclusion κ-good window which we call *κ-adaptive* and denote by $\Delta_\kappa(x_0)$.

The *adaptive estimate* of f at x_0 is defined as

$$\hat{f}(x_0) = \mathcal{S}^m_{\Delta_\kappa(x_0),x_0};$$

as we see, it depends on two "design" parameters, namely, on κ and m.

To find the adaptive window $\Delta_\kappa(x_0)$ one needs to perform only finite number of operations. Indeed, the quantities $\mathcal{S}^m_{\Delta,x_0}$ in fact depend only on the set M_Δ of observation points contained in Δ, so that only windows with different sets of observation points should be taken into account.

Our main goal is to evaluate the quality of the estimate $\hat{f}(\cdot)$. In this section we investigate the risk of the estimate at a fixed point x_0 under the following assumption on the signal f:

A. There exist $L > 0$, $l \ge 1$, $p \in [1,\infty]$ with

$$l \le m+1, \;\; pl > 1,$$

and an interval $\Delta_0 = [x_0 - \delta_0, x_0 + \delta_0] \subseteq [0,1]$ such that f is l times continuously differentiable on Δ_0 and

$$\left(\int_{\Delta_0} |f^{(l)}(t)|^p dt \right)^{1/p} \le L \tag{10}$$

(in the case of $p = \infty$ the left hand side of (10) is the usual L_∞-norm of a function on Δ_0).

We stress that the quantities δ_0, l, p, L were not involved into the construction of our estimate.

3.2 Main result

To evaluate quality of the adaptive estimate $\hat{f}(x_0)$, let us define the *ideal window* $\Delta^*_\kappa(x_0) = [x_0 - \delta^*_\kappa, x_0 + \delta^*_\kappa]$ by the following *balance equation*:

$$\delta^*_\kappa \equiv \delta^*_\kappa(x_0) = \max\left\{\delta \le \delta_0 : \frac{\kappa}{\sqrt{2n\delta}} \ge 2(1+\theta_m)\frac{\delta^{l-1/p}}{(l-1)!}\left(\int_{x_0-\delta}^{x_0+\delta} |f^{(l)}(x)|^p dx\right)^{1/p}\right\}. \quad (11)$$

Let us also define the set

$$\Xi \equiv \Xi(\kappa, m, n) = \left\{\omega \in \Omega : \sup_{\Delta \in \mathcal{W}, u \in \Delta} r^{-1}_{\Delta,u} |\sum_{x \in M_\Delta} \alpha_\Delta(x,u)\xi(x)| \le \kappa\right\}.$$

Note that the set Ξ is independent of f.

Theorem 1 *Let the assumption* **A** *hold with some l, p, L, Δ_0, $l \le m+1$, and let n be large enough, namely, let*

$$m+1 \le \min\left[\left(\frac{(l-1)!}{L(1+\theta_m)}\right)^{\frac{2p}{2pl+p-2}} (2n)^{\frac{2(pl-1)}{2pl+p-2}}; 2n\delta_0\right], \quad (12)$$

and let $\kappa \ge 1$. Then the error of the adaptive estimate $\hat{f}(x_0)$ associated with x_0, m, κ admits the following bound:

$$|\hat{f}(x_0) - f(x_0)| \le \begin{cases} \dfrac{6\theta_m \kappa}{\sqrt{2n\delta^*_\kappa(x_0)}}, & \omega \in \Xi \\ 4\theta_m\kappa + (1+\theta_m)\dfrac{\delta_0^{l-1/p} L}{(l-1)!} + \gamma_m \Theta, & \omega \in \Omega \backslash \Xi. \end{cases} \quad (13)$$

Here $\gamma_m > 0$ depends on m only and Θ is an appropriately chosen random variable which is a deterministic function of n, m and $\{\xi(x)\}_{x \in \Gamma_n}$ (so that Θ is independent of f, x_0 and the data involved into **A***) such that*

$$P\{\Theta > t\} \le n(n+1)(m+1)\sqrt{\frac{2}{\pi}}\int_t^\infty \exp\{-\frac{r^2}{2}\}dr. \quad (14)$$

Besides this, there exists constant $\pi_m > 0$ depending on m only such that

$$\Omega \backslash \Xi \subseteq \{\omega \in \Omega : \Theta > \pi_m \kappa\}. \quad (15)$$

Proof 1^0. Let us set

$$\omega_f(x,\delta) = \min_{p \in \mathcal{P}_m} \max_{z : |z-x| \le \delta} |f(z) - p(z)|,$$

where $\mathcal{P}_m$ is the space of polynomials of degrees $\le m$. Thus, $\omega_f(x,\delta)$ is the error of the best approximation of $f(x)$ by polynomials of degrees $\le m$ in the uniform norm.

The following well–known result yields an upper bound on the error of the estimator $\mathcal{S}^m_{\Delta,x_0}(y^n)$.

Lemma 1 *Let $\Delta \in \mathcal{D}(x_0)$. Then*

$$|\mathcal{S}^m_{\Delta,x_0}(y^n) - f(x_0)| \le (1+\theta_m)\omega_f(x_0,\delta) + |\sum_{x \in M_\Delta} \alpha_\Delta(x,x_0)\xi(x)|. \quad (16)$$

Proof Let $p(x)$ be the polynomial of degree $\leq m$ closest to $f(x)$ in the uniform norm on Δ. Then for the estimate (6)-(7) we have

$$|\mathcal{S}^m_{\Delta,x_0}(y^n) - f(x_0)| \leq |\sum_{x\in M_\Delta} \alpha_\Delta(x,x_0) f(x) - f(x_0)| + |\sum_{x\in M_\Delta} \alpha_\Delta(x,x_0)\xi(x)| \leq$$

$$|\sum_{x\in M_\Delta} \alpha_\Delta(x,x_0)(f(x)-p(x))| + |\sum_{x\in M_\Delta} \alpha_\Delta(x,x_0)p(x) - f(x_0)| + |\sum_{x\in M_\Delta} \alpha_\Delta(x,x_0)\xi(x)| \leq$$

$$\omega_f(x_0,\delta) r_{\Delta,x_0}\sqrt{N_\Delta} + |p(x_0) - f(x_0)| + |\sum_{x\in M_\Delta} \alpha_\Delta(x,x_0)\xi(x)| \leq$$

$$(1+\theta_m)\omega_f(x_0,\delta) + |\sum_{x\in M_\Delta} \alpha_\Delta(x)\xi(x)|.$$

Here the constraints of (7) were taken into account. □

Note that under assumption **A** for any window $\Delta \subseteq \Delta_0$ centered at x_0 one has

$$\omega_f(x_0,\delta) \leq \frac{\delta^{l-1/p}}{(l-1)!}\left(\int_{\Delta_0} |f^{(l)}(t)|^p dt\right)^{1/p}. \tag{17}$$

Relation (17) immediately follows from the Hölder inequality as applied to the integral formula for the reminder of the $(l-1)$th order Taylor expansion of f at x_0.

2^0. Let us verify that the ideal window belongs to $\mathcal{D}(x_0)$; by construction it is centered at x_0, therefore we need to verify only that it contains at least $m+1$ observation points. If $\delta^*_\kappa(x_0) < \delta_0$, then due to (10), (17) and (11) one has

$$\frac{\kappa}{\sqrt{2n\delta^*_\kappa}} \leq (\delta^*_\kappa)^{l-1/p}\frac{L}{(l-1)!}(1+\theta_m),$$

whence

$$\delta^*_\kappa \geq \delta_* \equiv \left(\frac{\kappa(l-1)!}{L\sqrt{2n}(1+\theta_m)}\right)^{\frac{2p}{2pl+p-2}}.$$

In the case of $\delta^*_\kappa \geq \delta_0$ we have $\delta^*_\kappa = \delta_0$; thus, we always have $\delta^*_\kappa \geq \min\{\delta_*,\delta_0\}$. By (12) both δ_* and δ_0 are not less than $(2n)^{-1}(m+1)$, so that at least $m+1$ points of the grid Γ_n do lie in Δ^*_κ. As a byproduct of our reasoning, we obtain the inequality $\delta^*_\kappa(x_0) \geq (2n)^{-1}(m+1)$, whence

$$2n\delta^*_\kappa(x_0) \leq N_{\Delta^*_\kappa(x_0)} \leq 4n\delta^*_\kappa(x_0). \tag{18}$$

3^0. We need the following simple

Lemma 2 *Let Δ be an admissible for x_0 window such that $\Delta \subseteq \Delta_\kappa(x_0)$ for some $\omega \in \Omega$. Then for this ω one has*

$$\left|\hat{f}(x_0) - f(x_0)\right| \leq 4\kappa r_{\Delta,x_0} + \left|\mathcal{S}^m_{\Delta,x_0} - f(x_0)\right|. \tag{19}$$

Proof By definition of the κ-adaptive window, the segment $\mathcal{I}_{\Delta_\kappa(x_0)}$ is nonempty; on the other hand, under the premise of the Lemma this segment is contained in both the segments

$$\left[\mathcal{S}^m_{\Delta_\kappa(x_0),x_0} - 2\kappa r_{\Delta_\kappa,x_0}, S^m_{\Delta_\kappa(x_0),x_0} + 2\kappa r_{\Delta_\kappa,x_0}\right]$$

and

$$\left[\mathcal{S}^m_{\Delta,x_0}-2\kappa r_{\Delta,x_0},\mathcal{S}^m_{\Delta,x_0}+2\kappa r_{\Delta,x_0}\right],$$

so that

$$|\mathcal{S}^m_{\Delta_\kappa(x_0),x_0}-\mathcal{S}^m_{\Delta,x_0}|\le 4\kappa r_{\Delta,x_0}$$

(we have taken into account that $r_{\Delta_\kappa(x_0),x_0}\le r_{\Delta,x_0}$ due to $\Delta\subseteq\Delta_\kappa(x_0)$), and (19) follows. □.

4^0. Let $\omega\in\Xi$. Then for any admissible Δ we have (Lemma 1 and the definition of Ξ)

$$|\mathcal{S}^m_{\Delta,x_0}-f(x_0)|\le(1+\theta_m)\omega_f(x_0,\delta)+\kappa r_{\Delta,x_0}.$$

By definition of the ideal window $\Delta^*_\kappa(x_0)$ and (17) one has

$$(1+\theta_m)\omega_f(x_0,\delta^*_\kappa(x_0))\le\frac{\kappa}{2\sqrt{2n\delta^*_\kappa(x_0)}}\le\frac{\kappa}{\sqrt{N_{\Delta^*_\kappa(x_0)}}}\le\kappa r_{\Delta^*_\kappa,x_0}$$

(the second inequality is given by (18), the third follows from the evident relation $r_{\Delta,u}\ge N_\Delta^{-1/2}$) and therefore

$$\Delta\in\mathcal{D}(x_0),\Delta\subseteq\Delta^*_\kappa(x_0)\ \Rightarrow|\mathcal{S}^m_{\Delta,x_0}-f(x_0)|\le 2\kappa r_{\Delta,x_0}. \tag{20}$$

Thus, all the segments

$$\left[\mathcal{S}^m_{\Delta,x_0}-2\kappa r_{\Delta,x_0},\mathcal{S}^m_{\Delta,x_0}+2\kappa r_{\Delta,x_0}\right]$$

associated with $\Delta\in\mathcal{D}(x_0)$, $\Delta\subseteq\Delta^*_\kappa(x_0)$, have a point in common, namely, $f(x_0)$. Hence, $\Delta^*_\kappa(x_0)$ is κ-good and therefore $\Delta^*_\kappa(x_0)\subseteq\Delta_\kappa(x_0)$. Thus, we have

$$\left|\hat f(x_0)-f(x_0)\right|\le 4\kappa r_{\Delta^*_\kappa(x_0),x_0}+|\mathcal{S}^m_{\Delta^*_\kappa(x_0),x_0}-f(x_0)|\le 6\kappa r_{\Delta^*_\kappa(x_0),x_0}\le$$

$$\le 6\theta_m\kappa N^{-1/2}_{\Delta^*_\kappa(x_0)}\le 6\theta_m\kappa(2n\delta^*_\kappa(x_0))^{-1/2}$$

(the first inequality is given by Lemma 2 applied with $\Delta=\Delta^*_\kappa(x_0)$, the second follows from (20), the third is (9), the fourth is given by (18)). The concluding inequality is that one required in (13).

5^0. Now consider the case of $\omega\in\Omega\backslash\Xi$. Due to (12), the ideal window $\Delta^*_\kappa(x_0)$ (which clearly belongs to Δ_0) is admissible for x_0; consequently, the minimal, with respect to inclusion, among the admissible windows, Δ_{min}, also is contained in Δ_0. By construction, $\hat f(x_0)=\mathcal{S}^m_{\Delta_\kappa(x_0),x_0}$, and $\Delta_{min}\subset\Delta_\kappa(x_0)$. Applying Lemma 2 to $\Delta=\Delta_{min}$, we get

$$|\hat f(x_0)-f(x_0)|\le 4\kappa r_{\Delta_{min},x_0}+|\mathcal{S}^m_{\Delta_{min},x_0}-f(x_0)|,$$

whence, by Lemma 1 and (9),

$$\left|\hat f(x_0)-f(x_0)\right|\le\frac{4\kappa\theta_m}{\sqrt{m+1}}+(1+\theta_m)\omega_f(x_0,\delta_{min})+\zeta\le$$

$$\le 4\theta_m\kappa+(1+\theta_m)\frac{\delta_0^{l-1/p}L}{(l-1)!}+\zeta,\ \ \zeta=\sup_{\Delta\in\mathcal{W},u\in\Delta}|\sum_{x\in M_\Delta}\alpha_\Delta(x,u)\xi(x)| \tag{21}$$

(we have taken into account that $\Delta_{min}\subset\Delta_0$). Note that ζ depends only on m,n and $\{\xi(x)\}_{x\in\Gamma_n}$. Our goal is to bound from above the random variable ζ.

$5^0.1$. Let us first assume that $m > 0$. It can be easily checked (see (7)) that $\alpha_\Delta(x,u)$, $\Delta \in \mathcal{W}$, is a polynomial of the degree m in u:

$$\alpha_\Delta(x,u) = \sum_{j=0}^{m} \alpha_\Delta^{(j)}(x) \left(\frac{u - a(\Delta)}{b(\Delta) - a(\Delta)} \right)^j, \tag{22}$$

where $a(\Delta)$ is the minimal, and $b(\Delta)$ is the maximal observation point in Δ. In addition, it is known that $|\alpha_\Delta(x,u)| \le \theta_m/N_\Delta$, $u \in \Delta$ (cf. (8)).

By sequential applying the Markov inequality to the polynomial $\alpha_\Delta(x,\cdot)$ we easily obtain that there exists a constant τ_m depending on m only such that for any j

$$|\alpha_\Delta^{(j)}(x)| \le \tau_m \max_{u\in\Delta} |\alpha_\Delta(x,u)| \le \tau_m \theta_m / N_\Delta. \tag{23}$$

It implies that

$$\sigma_{j,\Delta} \equiv \left(\sum_{x \in M_\Delta} |\alpha_\Delta^{(j)}(x)|^2 \right)^{1/2} \le \tau_m \theta_m / \sqrt{N_\Delta} \le \tau_m \theta_m r_{\Delta,u}, \; u \in \Delta \tag{24}$$

(the concluding inequality follows from the evident relation $r_{\Delta,u} \ge 1/\sqrt{N_\Delta}$). Thus, one has

$$\begin{aligned} |\sum_{x\in M_\Delta} \alpha_\Delta(x,u)\xi(x)| = |\sum_{j=0}^{m} \left(\frac{u - a(\Delta)}{b(\Delta) - a(\Delta)} \right)^j \sum_{x\in M_\Delta} \alpha_\Delta^{(j)}(x)\xi(x)| \le \\ 2^m(m+1) \max_{0\le j\le m} |\sum_{x\in M_\Delta} \alpha_\Delta^{(j)}(x)\xi(x)| \end{aligned} \tag{25}$$

(we have taken into account that if $m > 0$, $\Delta \in \mathcal{W}$ and $u \in \Delta$, then $|(u - a(\Delta))/(b(\Delta) - a(\Delta))| \le 2$). The resulting inequality evidently holds in the case of $m = 0$ as well (where $\alpha_\Delta(x,u) = \alpha_\Delta^{(0)}(x)$ are independent of u).

Comparing (21) and (25), we see that for any $m \ge 0$ one has

$$\zeta \le 2^m(m+1) \sup_{\Delta\in\mathcal{W}} \max_{0\le j\le m} |\sum_{x\in M_\Delta} \alpha_\Delta^{(j)}(x)\xi(x)|. \tag{26}$$

$5^0.2$. Observe that any two of *equivalent* windows $\Delta, \Delta' \in \mathcal{W}$, i.e. those with identical sets of the observation points, yield the same random variable in the right hand side of (26). Therefore we can rewrite (26) as

$$\zeta \le 2^m(m+1) \max_{\Delta\in\mathcal{W}^*} \max_{0\le j\le m} |\sum_{x\in M_\Delta} \alpha_\Delta^{(j)}(x)\xi(x)|, \tag{27}$$

$\mathcal{W}^*$ being a maximal, with respect to inclusion, family of *mutually non-equivalent* windows from $\mathcal{W}$. The set $\mathcal{W}^*$ clearly is finite of cardinality $\le n(n+1)$.

Now let

$$\eta_{j,\Delta} = \sigma_{j,\Delta}^{-1} \sum_{x\in M_\Delta} \alpha_\Delta^{(j)}(x)\xi(x),$$

($\sigma_{j,\Delta}$ is given by (24)), so that $\eta_{j,\Delta}$ are $\mathcal{N}(0,1)$ random variables, and let

$$\Theta = \max_{\Delta\in\mathcal{W}^*} \max_{0\le j\le m} |\eta_{j,\Delta}|.$$

It is immediately seen that Θ satisfies (14).

6^0. Let us verify that with the indicated choice of Θ the second relation in (13) holds. Indeed, from (9) and (24) we know that

$$\sigma_{j,\Delta} \le \tau_m \theta_m^2, \tag{28}$$

so that (27) results in

$$\zeta \le 2^m(m+1)\tau_m\theta_m^2\Theta \equiv \gamma_m\Theta,$$

which combined with (21) leads to the desired inequality.

7^0. To complete the proof, we should verify (15). By virtue of (25), for $\Delta \in \mathcal{W}$ and $u \in \Delta$ one has

$$|\sum_{x\in M_\Delta} \alpha_\Delta(x,u)\xi(x)| \le 2^m(m+1)\max_{0\le j\le m} |\sum_{x\in M_\Delta} \alpha_\Delta^{(j)}(x)\xi(x)|,$$

so that

$$\{\omega\in\Omega : r_{\Delta,u}^{-1}|\sum_{x\in M_\Delta}\alpha_\Delta(x,u)\xi(x)| > \kappa\} \subseteq$$

$$\bigcup_{0\le j\le m}\{\omega\in\Omega : 2^m(m+1)r_{\Delta,u}^{-1}|\sum_{x\in M_\Delta}\alpha_\Delta^{(j)}(x)\xi(x)| > \kappa\} =$$

$$\bigcup_{0\le j\le m}\{\omega\in\Omega : \sigma_{j,\Delta}^{-1}|\sum_{x\in M_\Delta}\alpha_\Delta^{(j)}(x)\xi(x)| > \frac{\kappa r_{\Delta,u}}{2^m(m+1)\sigma_{j,\Delta}}\} \subseteq$$

$$\bigcup_{0\le j\le m}\{\omega\in\Omega : |\eta_{j,\Delta}| > \frac{\kappa}{2^m(m+1)\theta_m^2\tau_m}\}.$$

Here we have taken into account the definition of $\eta_{j,\Delta}$ and (28). Consequently,

$$\Omega\backslash\Xi = \bigcup_{\Delta\in\mathcal{W},u\in\Delta}\left\{\omega\in\Omega : r_{\Delta,u}^{-1}|\sum_{x\in M_\Delta}\alpha_\Delta(x,u)\xi(x)| > \kappa\right\} \subseteq$$

$$\bigcup_{\Delta\in\mathcal{W},0\le j\le m}\left\{\omega\in\Omega : |\eta_{j,\Delta}| > \frac{\kappa}{2^m(m+1)\theta_m^2\tau_m}\right\}.$$

As before, the union over $\Delta \in \mathcal{W}, 0 \le j \le m$ in the right hand side can be replaced with the union over $\Delta \in \mathcal{W}^*, 0 \le j \le m$, and we come to

$$\Omega\backslash\Xi \subseteq \left\{\omega\in\Omega : \Theta > \frac{\kappa}{2^m(m+1)\theta_m^2\tau_m}\right\},$$

as required in (15). ∎.

Theorem 1 allows to get upper bounds on the local risk.

Corollary 1 *There exists constant c_m depending on m such that the estimate $\hat f(x_0)$ associated with*

$$\kappa = c_m\sqrt{\ln n} \tag{29}$$

under assumptions of Theorem 1 satisfies the relation

$$\left(\mathcal{E}|\hat f(x_0) - f(x_0)|^2\right)^{1/2} \le O(1)\left[\left(\frac{\ln n}{n}\right)^{\frac{pl-1}{2pl+p-2}} L^{\frac{p}{2pl+p-2}} + \sqrt{\frac{\ln n}{n\delta_0}}\right] \tag{30}$$

(here and in what follows all $O(1)$'s depend on m only).

Proof From (14) it immediately follows that

$$\mathcal{E}\Theta^4 \le O(1)(\ln n)^2.$$

Under appropriate choice of (depending on m only) constant c_m we have

$$P\{\Theta > \pi_m c_m \sqrt{\ln n}\} \le O(1)n^{-2(m+1)}.$$

With this choice of c_m and with κ given by (29) one has

$$\left[\mathcal{E}\left(\left[4\kappa\theta_m + \frac{\delta_0^{l-1/p}L}{(l-1)!} + \gamma_m\Theta\right]\chi_{\{\Theta>\pi_m\kappa\}}\right)^2\right]^{1/2} \le O(1)\left[\kappa + \delta_0^{l-1/p}L\right](P\{\Theta > \pi_m\kappa\})^{1/2} +$$

$$O(1)\left((\mathcal{E}\Theta^4)^{1/2}(P\{\Theta > \pi_m\kappa\})^{1/2}\right)^{1/2} \le O(1)Ln^{-(m+1)} + O(1)n^{-(m+1)/2}\sqrt{\ln n}.$$

(from now on χ_A is the characteristic function of an event $A \subset \Omega$). Observe that due to (17) and (11) one has

$$\frac{\kappa}{\sqrt{2n\delta_\kappa^*(x_0)}} \le O(1)\left[\left(\frac{\ln n}{n}\right)^{\frac{pl-1}{2pl+p-2}} L^{\frac{p}{2pl+p-2}} + \sqrt{\frac{\ln n}{n\delta_0}}\right].$$

Therefore, (13) and (15) imply

$$\left(\mathcal{E}|\hat{f}(x_0) - f(x_0)|^2\right)^{1/2} \le O(1)\left[\left(\frac{\ln n}{n}\right)^{\frac{pl-1}{2pl+p-2}} L^{\frac{p}{2pl+p-2}} + \sqrt{\frac{\ln n}{n\delta_0}} + Ln^{-(m+1)}\right]. \qquad (31)$$

It is straightforward to verify that one has

$$Ln^{-(m+1)} < O(1)\left(\frac{\ln n}{n}\right)^{\frac{l}{2l+1}} L^{\frac{1}{2l+1}}. \qquad (32)$$

Indeed, it follows from (12) that $L < O(1)n^{l-1/p}\sqrt{\ln n}$ with a constant $O(1)$ depending on m only. Therefore, for (32) it suffices to show that

$$\left(\frac{\ln n}{n}\right)^{\frac{l}{2l+1}} n^{m+1} > O(1)(\ln n)^{\frac{l}{2l+1}} n^{\frac{2l(pl-1)}{2l+1}}.$$

This inequality is true due to $l \le m+1$. Thus, (32) combined with (31) results in the concluding inequality in (30). ∎.

Note that the quality of the adaptive estimate given by the Corollary 1 coincides, up to a logarithmic in n factor, with the lower bound on the minimax risk corresponding to the case when the smoothness of the function is specified completely. It is known ([5]) that for estimating a function of unknown smoothness at a fixed point (as our estimate does) the indicated factor, in a sense, cannot be eliminated. Thus, our adaptive estimate adapts, in the optimal way, to the unknown smoothness and is spatial adaptive. Moreover, we show that the estimate behaves "well" for estimating the whole function $f(x)$ over wide range of global risks.

4 Global risks

Now we address the issue of evaluating the global risks of our estimate. We again assume that f satisfies **A** with respect to some $x_0 \in (0,1)$; our goal is to evaluate the risk

$$\mathcal{R}_{q,\Delta;1}(\hat{f},f) = \mathcal{E}\|\hat{f}(\cdot) - f(\cdot)\|_{q,\Delta}$$

of the estimate associated with a segment Δ which is strictly inside Δ_0 (we restrict ourselves with the segments belonging to the interior of Δ_0 in order to avoid "boundary effects" which clearly badly influence our "symmetric" estimator). For the sake of simplicity, let us set

$$\Delta = \Delta(\Delta_0) = [x_0 - \delta_0/2, x_0 + \delta_0/2]$$

(recall that $\Delta_0 = [x_0 - \delta_0, x_0 + \delta_0]$); thus, we are interested in global risk of our estimate on the twice smaller than Δ_0 concentric to Δ_0 segment. The answer is given by the following

Theorem 2 *Let the assumption* **A** *hold with some* $\Delta_0 = [x_0 - \delta_0, x_0 + \delta_0]$, $p \in [1,\infty]$, $l \le m+1$ *and* $L > 0$, *and let* n *be large enough, namely, let*

$$m+1 \le \min\left[\left(\frac{(l-1)!}{L(1+\theta_m)}\right)^{\frac{2p}{2pl+p-2}}(2n)^{\frac{2(pl-1)}{2pl+p-2}}; \frac{1}{2}n\delta_0\right]. \tag{33}$$

Then for the estimate $\hat{f}(\cdot)$ *associated with the given* m, n *and some* $\kappa \ge 1$ *one has*

(i) *On the set* $\Xi = \Xi(\kappa, m, n)$

$$\|\hat{f}(\cdot) - f(\cdot)\|_{q,\Delta} \le O(1)\left[\delta_0^{\frac{1}{q}-\frac{1}{2}}\frac{\kappa}{\sqrt{n}} + \delta_0^{\psi}L^{\varphi}\left(\frac{\kappa^2}{n}\right)^{\sigma}\right], \tag{34}$$

where

- *in the case* $q \ge p(2l+1)$: $\psi = 0$, $\varphi = \dfrac{1-2/q}{2l+1-2/p}$, $\sigma = \dfrac{l-1/p+1/q}{2l+1-2/p}$;
- *in the case* $q \le p(2l+1)$: $\psi = \dfrac{1}{q} - \dfrac{1}{p(2l+1)}$, $\varphi = \dfrac{1}{2l+1}$, $\sigma = \dfrac{l}{2l+1}$.

Here and from now on $O(1)$*'s are positive quantities depending on* m *only.*

(ii) *On the set* $\Omega\backslash\Xi$

$$\|\hat{f}(\cdot) - f(\cdot)\|_{q,\Delta} \le O(1)[\kappa + L\delta_0^{l-1/p} + \Theta], \tag{35}$$

Θ *being the same random variable as in Theorem 1.*

Proof To avoid extra comments, in the below proof we restrict ourselves with the case of $q < \infty$; it is immediately seen from the proof that our reasoning allows passing to limit as $q \to \infty$, thus leading to the announced relations for the case of $q = \infty$.

1^0. Let f satisfy assumption **A**. Consider first the case of $p < \infty$.

Denote

$$\Delta_0 = [u_0, v_0];\ \Delta = [u, v];\ \zeta = \kappa^2 n^{-1};\ \beta = \frac{p}{2pl+p-2}.$$

Given point $x \in \Delta' \equiv [\frac{1}{2}(u+u_0), \frac{1}{2}(v+v_0)]$ let us define $\delta^*_+(x)$ as the largest $\delta \le \delta_0/4$ such that

$$\delta \le (\eta^2\zeta)^\beta \left[\int_x^{x+\delta} |f^{(l)}(t)|^p dt\right]^{-2\beta/p}, \quad \eta \equiv \frac{(l-1)!}{2\sqrt{2}(1+\theta_m)}.$$

Denote

$$R_+(x,\delta) = \int_x^{x+\delta} |f^{(l)}(t)|^p dt;$$

then

$$\delta^*_+(x) \le (\eta^2\zeta)^\beta \left[R_+(x, \delta^*_+(x))\right]^{-2\beta/p}, \tag{36}$$

with the inequality being equality for *proper* x, i.e. for those with $\delta^*_+(x) < \delta_0/4$. Observe that

$$R_+(x, \delta^*_+(x)) \le \int_{\Delta_0} |f^{(l)}(t)|^p dt \le L^p \tag{37}$$

due to assumption **A**.

2^0. Let us set $u_1 = \frac{1}{2}(u_0+u)$, $u_2 = u_1 + \delta^*_+(u_1)$, $u_3 = u_2 + \delta^*_+(u_2), \ldots$; we terminate this construction at the step N, when it turns out that $\frac{1}{2}(u_N + u_{N+1}) \ge v$. Due to (37) and the fact that (36) is an equality for the proper x, we have that the quantities $\delta^*_+(x)$ are bounded away from zero for all $x \in \Delta'$; whence the above N is well defined. Denote $U_i = [u_i, u_{i+1}]$, $i = 1, \ldots, N$ and let x_i and $2d_i$ be the midpoint of the segment U_i and its length respectively. For $1 \le i \le N$ define

$$\rho_i = \rho_i(u_i) = \int_{u_i}^{u_{i+1}} |f^{(l)}(t)|^p dt.$$

By construction and due to (10) one has

$$\sum_{i=1}^N \rho_i \le L^p. \tag{38}$$

Let $X_i = [x_i, x_{i+1}]$, $i = 1, \ldots, N$; then, clearly,

$$\Delta \subset \bigcup_{i=1}^{N-1} X_i. \tag{39}$$

3^0. We need the following simple

Lemma 3 *Let $d^*_i = \min\{d_i, d_{i+1}\}$, $i = 1, \ldots, N$; then for any $x \in X_i$ one has*

$$\delta^*_+(x) \ge \max\{d^*_i, |x - u_{i+1}|\}. \tag{40}$$

Proof There are two possible cases: $x_i \le x \le u_{i+1}$ and $u_{i+1} \le x \le x_{i+1}$. We shall prove the statement for the first case only; the second one is completely symmetric.

To prove the required inequality it suffices to consider the case of proper x. Indeed, if $\delta^*_+(x) = \delta_0/4$ then the statement is fulfilled trivially by definition of d_i.

1) Consider the case of

$$d^*_i \le |x - u_{i+1}|.$$

In contrast, assume that the opposite to (40) inequality holds. Then, evidently

$$[x, x+\delta^*_+(x)] \subseteq [x, u_{i+1}] \subset [u_i, u_{i+1}]. \tag{41}$$

It follows from (36) that

$$(\delta_+^*(x))^{-p/(2\beta)}(\eta^2\zeta)^{p/2} = R_+(x,\delta_+^*(x)) \le \rho_i \le (2d_i)^{-p/(2\beta)}(\eta^2\zeta)^{p/2};$$

whence $\delta_+^*(x) \ge 2d_i$ and this contradicts (41).

2) Now consider the case

$$d_i^* > |x - u_{i+1}|.$$

In this case

$$(\delta_+^*(x))^{-p/(2\beta)}(\eta^2\zeta)^{p/2} = R_+(x,\delta_+^*(x)) \le \rho_i + \rho_{i+1} \le 2(2d_i^*)^{-p/(2\beta)}(\eta^2\zeta)^{p/2},$$

therefore

$$\delta_+^*(x) \ge 2^{1-(2\beta)/p} d_i^* \ge d_i^*,$$

as claimed (here we have taken into account that $0 \le 2\beta/p \le 1$). □.

4^0. Let $x \in X_i$, $i = 1,\dots,N$. Now let us define $\delta_\kappa^*(x)$ as the largest $\delta \le \delta_0/4$ such that (cf. (11), (36))

$$\delta \le (\eta^2\zeta)^\beta [R(x,\delta)]^{-2\beta/p}, \quad R(x,\delta) = \int_{x-\delta}^{x+\delta} |f^{(l)}(t)|^p dt.$$

Note that (33) and **A** ensure the assumptions of Theorem 1. Therefore applying Theorem 1 to the estimate $\hat f(x)$ and the "smoothness interval" $\tilde\Delta(x) = [x - \delta_0/4, x + \delta_0/4]$ (so that one should substitute into formulation of the theorem x instead of x_0 and $\tilde\Delta(x)$ instead of Δ_0) we conclude that

$$|\hat f(x) - f(x)| \le \begin{cases} \dfrac{6\kappa\theta_m}{\sqrt{2n\delta_\kappa^*(x)}}, & \omega \in \Xi \\ 4\kappa\theta_m + (1+\theta_m)\dfrac{(\delta_0/4)^{l-1/p}L}{(l-1)!} + \gamma_m\Theta, & \omega \in \Omega\backslash\Xi \end{cases}$$

Now observe that

$$R(x,\delta) \le R_+(x,\delta),$$

so that for each x with $\delta_\kappa^*(x) \le \delta_0/4$ we have

$$\delta_+^*(x) \ge \delta_\kappa^*(x); \tag{42}$$

if $\delta_\kappa^*(x) = \delta_0/4$, then, evidently, $\delta_+^*(x) = \delta_0/4$, whence (42) in fact holds for any x. Therefore we obtain that

$$|\hat f(x) - f(x)| \le \begin{cases} \dfrac{6\kappa\theta_m}{\sqrt{2n\delta_+^*(x)}}, & \omega \in \Xi \\ 4\kappa\theta_m + (1+\theta_m)\dfrac{(\delta_0/4)^{l-1/p}L}{(l-1)!} + \gamma_m\Theta, & \omega \in \Omega\backslash\Xi \end{cases} \tag{43}$$

4^0.1. First assume that $\omega \in \Xi$; then due to (39) and (43) we have

$$\|\hat f(x) - f(x)\|_{q,\Delta}^q \le \left[\frac{6\theta_m}{\sqrt{2}}\right]^q \sum_{i=1}^{N-1} \mathcal{J}_i, \tag{44}$$

where

$$\mathcal{J}_i = \zeta^{q/2} \int_{X_i} (\delta_+^*(x))^{-q/2} dx.$$

Now, let I' be the set of those indexes $i = 1, \ldots, N-1$ for which both the segments U_i and U_{i+1} are of the length $\delta_0/4$ (i.e. u_i and u_{i+1} are not proper), and let I'' be the set of the remaining indexes. If $i \in I'$ and $x \in X_i$ then we obtain due to (33) that

$$\mathcal{J}_i \leq C_1^q \zeta^{q/2} \delta_0^{1-q/2} \tag{45}$$

(from now one C_i stand for the constant depending only on m). Furthermore, it is evident that cardinality of I' does not exceed $O(1)$, whence

$$\mathcal{I}_{1,q} \equiv \left[\sum_{i \in I'} \mathcal{J}_i \right]^{1/q} \leq C_2 \zeta^{1/2} \delta_0^{\frac{1}{q} - \frac{1}{2}}. \tag{46}$$

Now assume that $i \in I''$ and

$$q \geq 2;$$

let also, for the sake of definiteness, $d_i^* = d_i \leq d_{i+1}$ (the opposite case is completely symmetric). Applying Lemma 3 we conclude that

$$\begin{aligned} \zeta^{q/2} \int_{X_i} (\delta_+^*(x))^{-q/2} dx \quad &\leq \zeta^{q/2} \left(\int_{x_i}^{u_{i+1}+d_i} (d_i^*)^{-q/2} dx + \int_{d_i}^{d_{i+1}} r^{-q/2} dr \right) \leq \\ &\leq C_3^q \zeta^{q/2} (d_i^*)^{1-q/2}. \end{aligned} \tag{47}$$

Denote

$$\hat{\rho}_i \equiv \max\{\rho_i, \rho_{i+1}\} \leq \int_{U_i \cup U_{i+1}} |f^{(l)}(t)|^p dt.$$

Since one of the segments U_i and U_{i+1} is of the length $< \delta_0/4$, the smaller of the segments, say $U_{i'}$, also is of the length $< \delta_0/4$; by construction of the segments U_i it is possible only if

$$2d_i^* = |D_{i'}| = (\eta^2 \zeta)^\beta (\rho_{i'})^{-2\beta/p} \geq (\eta^2 \zeta)^\beta (\hat{\rho}_i)^{-2\beta/p}.$$

In view of (47) and $q \geq 2$, we have

$$\zeta^{q/2} \int_{X_i} (\delta_+^*(x))^{-q/2} dx \leq C_4^q \zeta^{(q - \beta(q-2))/2} \hat{\rho}_i^{\beta(q-2)/p}.$$

Whence

$$\mathcal{I}_{2,q}^q \equiv \sum_{i \in I''} \mathcal{J}_i \leq C_4^q \zeta^{(q-\beta(q-2))/2} \sum_{i \in I''} \hat{\rho}_i^{\beta(q-2)/p}. \tag{48}$$

Further, due to (38) one has

$$\sum_{i=1}^{N-1} \hat{\rho}_i \leq 2L^p. \tag{49}$$

In the case $q \geq p(2l+1)$ we have $\beta(q-2)/p \geq 1$, so that (48), (49) imply

$$\mathcal{I}_{2,q}^q \leq C_5^q \zeta^{(q-\beta(q-2))/2} L^{\beta(q-2)},$$

whence, substituting $\beta = p/(2pl+p-2)$, we come to

$$q \geq p(2l+1) \Rightarrow \mathcal{I}_{2,q} \leq C_5 \zeta^{\frac{l-1/p+1/q}{2l+1-2/p}} L^{\frac{1-2/q}{2l+1-2/p}}. \tag{50}$$

Combining (46) and (50), we come to the bound announced in (34) for the case of $q \geq p(2l+1)$.

Now consider the case of $q \leq p(2l+1)$. From the Hölder inequality it follows that

$$\mathcal{I}_{2,q} \leq \delta_0^{\frac{1}{q}-\frac{1}{q'}} \mathcal{I}_{2,q'}, \quad 1 \leq q \leq q'.$$

Setting $q' = p(2l+1)$, we obtain

$$1 \leq q \leq p(2l+1) \Rightarrow \mathcal{I}_{2,q} \leq C_6 \delta_0^{\frac{1}{q}-\frac{1}{p(2l+1)}} L^{1/(2l+1)} \zeta^{l/(2l+1)};$$

this bound together with (46) lead to (34) for the case $q \leq p(2l+1)$.

4^0.2. We have established (34) for the case $p < \infty$. To get the result for the case $p = \infty$ it suffices to pass in (34) to limit as p tends to ∞.

5^0. The statement announced in (ii) is an immediate consequence of (43). ∎.

Corollary 2 *Let the estimate $\hat{f}(\cdot)$ be associated with the choice*

$$\kappa = c_m \sqrt{\ln n},$$

where c_m is specified in Corollary 1. Then under conditions of Theorem 2 one has

$$\left(\mathcal{E}\|\hat{f}(\cdot) - f(\cdot)\|^2_{q,\Delta}\right)^{1/2} \leq O(1) \left[\delta_0^{\frac{1}{q}-\frac{1}{2}} \left(\frac{\ln n}{n}\right)^{1/2} + \delta_0^{\psi} L^{\varphi} \left(\frac{\ln n}{n}\right)^{\sigma}\right], \tag{51}$$

where the constants ψ, φ, and σ are specified in Theorem 2.

Proof of the corollary word by word follows that one of Corollary 1, with Theorem 2 playing the role of Theorem 1.

Let us compare the bounds on global (on Δ) risk of our estimate given by Corollary 2 with the *optimal* minimax risk

$$\mathcal{R}^*_{q,\Delta;2}(n, \mathcal{F})$$

corresponding to the family

$$\mathcal{F} = \{f : f \textit{ is } l \textit{ times continuously differentiable on } \Delta_0, \|f^{(l)}\|_{p,\Delta_0} \leq L\}$$

of all signals satisfying the assumption **A**. Consider the case of large n; then (33) is valid, and the quality of our estimate is given by (51). As $n \to \infty$, the right hand side of (51) is equal, up to a constant factor C depending on m, p, q, δ_0 and L, to

- $\phi(n) = \left(\frac{\ln n}{n}\right)^{\frac{l-1/p+1/q}{2l+1-2/p}}$, if $q \geq p(2l+1)$ (case I, "strong" accuracy measure);
- $\phi(n) = (\ln n)^{\frac{l}{2l+1}} n^{-l/(2l+1)}$, if $q \leq p(2l+1)$ (case II, "weak" accuracy measure).

It is known ([8]) that in Case I the optimal risk $\mathcal{R}^*$ behaves itself exactly as $\phi(\cdot)$, so that in this case our adaptive estimate is optimal in order. It is well-known that in the remaining case the optimal risk $\mathcal{R}^*$ is of order of $n^{-l/(2l+1)}$, so that here our estimate is "near-optimal": $\phi(\cdot)$ coincides with $\mathcal{R}^*$ within a logarithmic in n factor. Thus, our estimate indeed is spatial adaptive.

5 Estimation of the derivatives

Here we are interested in restoring the derivative $f^{(s)}(\cdot)$ via observations (1). It turns out that our estimate can be easily adapted to solve this problem. For reasons of space we describe construction of the estimate and present main results omitting the proofs.

As before, let us fix positive integer $m \geq s$ and let $\mathcal{S}^m_{\Delta,u}(y^n)$ be the least square estimator of the order m associated with the window Δ and the point $u \in \Delta$. Assume that $\{\alpha_\Delta(x,u)\}_{x\in M_\Delta}$ form the solution of (7). Estimates for the derivatives can be obtained from $\mathcal{S}^m_{\Delta,u}(y^n)$ by sequential differentiation. Formally, let

$$\mathcal{S}^{m,s}_{\Delta,u}(y^n) = \sum_{x\in M_\Delta} \beta_{\Delta,s}(x,u)y(x),$$

where $\beta_{\Delta,s}(x,u) = \partial^s\alpha_\Delta(x,u)/\partial u^s$ (in this notation $\mathcal{S}^{m,0}_{\Delta,u}(y^n) \equiv \mathcal{S}^m_{\Delta,u}(y^n)$ and $\beta_{\Delta,0}(x,u) \equiv \alpha_\Delta(x,u)$). Observe that $\{\beta_{\Delta,s}(x,u)\}_{x\in M_\Delta}$ form solution to the following optimization problem

$$\begin{aligned} &\sum_{x\in M_\Delta} \beta^2_{\Delta,s}(x,u) &=&\ min \\ \textit{subject to}\quad &\sum_{x\in M_\Delta} \beta_{\Delta,s}(x,u)x^j &=&\ 0, \quad j = 0,\dots,s-1 \\ &\sum_{x\in M_\Delta} \beta_{\Delta,s}(x,u)x^j &=&\ \frac{j!}{(j-s)!}u^{j-s}, \quad j = s,\dots,m. \end{aligned}$$

Hence it follows that the estimate $\mathcal{S}^{m,s}_{\Delta,u}(y^n)$ is well–defined for any window Δ containing at least $m-s+1$ observation points. As before, such windows centered at u are called *admissible* for $u \in (0,1)$ and set of all admissible for u windows is designated by $\mathcal{D}_s(u)$. Similarly, *the index* of the admissible window Δ at u is defined as

$$r_{\Delta,u,s} = \left(\sum_{x\in M_\Delta} \beta^2_{\Delta,s}(x,u)\right)^{1/2}.$$

In view of (22) and (23) we have

$$|\beta_{\Delta,s}(x,u)| \leq \frac{\vartheta_{m,s}}{\delta^s N_\Delta}, \quad x \in M_\Delta,$$

where $\vartheta_{m,s} = m\tau_m\theta_m 2^{m-s+1}m!/s!$. Consequently,

$$r_{\Delta,u,s} \leq \frac{\vartheta_{m,s}}{\delta^s\sqrt{N_\Delta}}.$$

Define κ*–good* window $\Delta \in \mathcal{D}_s(x_0)$ according to Definition 2 replacing $\mathcal{S}^m_{\Delta,x_0}$, r_{Δ,x_0} by $\mathcal{S}^{m,s}_{\Delta,x_0}$ and $r_{\Delta,x_0,s}$ respectively. Maximal with respect to inclusion κ-good window is called κ*–adaptive* and is denoted by $\Delta_\kappa(x_0)$. The *adaptive estimate* of $f^{(s)}$ at x_0 is defined as

$$\hat{f}^{(s)}(x_0) = \mathcal{S}^{m,s}_{\Delta_\kappa,x_0}.$$

It turns out that the results similar to Theorems 1, 2 and Corollaries 1, 2 can be established for the adaptive estimate $\hat{f}^{(s)}(x_0)$. We present them in the following two theorems.

Theorem 3 *Let assumption* **A** *hold with some* l, p, L, Δ_0, $s+1 \le l \le m+1$, $p(l-s) > 1$, *and let the sample size* n *be large enough, so that*

$$m - s + 1 \le \min\left[\left(\frac{1}{L(1+\vartheta_{m,s})}\right)^{\frac{2p}{2pl+p-2}}(2n)^{\frac{2(pl-1)}{2pl+p-2}};\ 2n\delta_0\right].$$

Then there exists constant $c_{m,s}$ *depending on* m *and* s *only so that for the estimate* $\hat{f}^{(s)}(x_0)$ *associated with the choice*

$$\kappa = c_{m,s}\sqrt{\ln n} \tag{52}$$

one has

$$\left(\mathcal{E}|\hat{f}^{(s)}(x_0) - f^{(s)}(x_0)|^2\right)^{1/2} \le O(1)\left[\left(\frac{\ln n}{n}\right)^{\frac{p(l-s)-1}{2pl+p-2}} L^{\frac{p(2s+1)}{2pl+p-2}} + \sqrt{\frac{\ln n}{n}}\delta_0^{-s-1/2}\right], \tag{53}$$

where the constant $O(1)$ *depends on* m *and* s *only.*

Thus, the upper bound (53) on the quality of the adaptive estimate $\hat{f}^{(s)}(x_0)$ coincides up to a ln n–factor with the lower bound on the minimax risk in the case of known smoothness.

Now our goal is to evaluate the global risk

$$\mathcal{R}^s_{q,\Delta;1}(\hat{f}, f) = \mathcal{E}\|\hat{f}^{(s)}(\cdot) - f^{(s)}(\cdot)\|_{q,\Delta},$$

where as before

$$\Delta = \Delta(\Delta_0) = [x_0 - \delta_0/2, x_0 + \delta_0/2].$$

Theorem 4 *Let assumption* **A** *hold with some* $\Delta_0 = [x_0 - \delta_0, x_0 + \delta_0]$, $p \in [1, \infty]$, $s+1 \le l \le m+1$, $p(l-s) > 1$, *and* $L > 0$ *and let* n *be large enough so that*

$$m - s + 1 \le \min\left[\left(\frac{1}{L(1+\vartheta_{m,s})}\right)^{\frac{2p}{2pl+p-2}}(2n)^{\frac{2(pl-1)}{2pl+p-2}};\ \frac{1}{2}n\delta_0\right].$$

Then for the adaptive estimate $\hat{f}^{(s)}(\cdot)$ *associated with the choice (52) one has*

$$\left(\mathcal{E}\|\hat{f}^{(s)}(\cdot) - f^{(s)}(\cdot)\|^2_{q,\Delta}\right)^{1/2} \le O(1)\left[\delta_0^{\frac{1}{q}-s-\frac{1}{2}}\left(\frac{\ln n}{n}\right)^{1/2} + \delta_0^{\psi}L^{\varphi}\left(\frac{\ln n}{n}\right)^{\sigma}\right],$$

where

- *in the case of* $q \ge p\dfrac{2l+1}{2s+1}$: $\psi = 0$, $\varphi = \dfrac{2s+1-2/q}{2l+1-2/p}$, $\sigma = \dfrac{l-s-1/p+1/q}{2l+1-2/p}$;
- *in the case of* $q \le p\dfrac{2l+1}{2s+1}$: $\psi = \dfrac{1}{q} - \dfrac{2s+1}{p(2l+1)}$, $\varphi = \dfrac{2s+1}{2l+1}$, $\sigma = \dfrac{l-s}{2l+1}$.

Here $O(1)$ *is positive constant depending on* m *and* s *only.*

The proofs of Theorems 3, 4 are similar to those of Theorems 1 and 2. Comparing the upper bound given by Theorem 4 with the lower bound on the minimax risk for estimation of the derivatives (cf. [8]) we conclude that in the case of $q > p(2l+1)/(2s+1)$ our adaptive estimate is optimal in order. In the remaining case it is "near–optimal" within a logarithmic in n factor.

6 Spatial adaptivity: limits of performance

Here we demonstrate that our estimate possesses, in a sense, the best possible abilities for spatial adaptation. To pose the question formally, let us act as follows.

Let $D \subset [0,1]$ be a segment, L be positive real, l be positive integer, and $p \in [1,\infty]$. Let $W_p^l(D,L)$ denote the standard Sobolev class

$$W_p^l(D,L) = \{f \in \mathrm{C}^l \mid \|f^{(l)}\|_{p,D} \le L\}.$$

Let D' be the concentric to D and smaller than D in a once for ever fixed ratio (say, by 50%) segment. For $q \in [1,\infty)$ and a positive integer n, let, as in Introduction,

$$\mathcal{R}^*_{q,D',q}(n; W_p^l(D,L)) = \inf_{\tilde{f}\in F_n} \sup_{f\in W_p^l(D,L)} \left[\mathcal{E}\{\|\tilde{f} - f\|^q_{q,D'}\}\right]^{1/q}$$

be the minimax risk, measured in the norm $\|\cdot\|_{q,D'}$, of estimating a signal $f \in W_p^l(D,L)$ via observations (1); F_n stands here for the family of all estimates with n observation points on $[0,1]$. Now let

$$\bar{f} \equiv \{\bar{f}_n \in F_n\}_{n=1}^{\infty}$$

be an estimation method, and let

$$\mathcal{R}_{q,D',q}(\bar{f}_n; W_p^l(D,L)) = \sup_{f\in W_p^l(D,L)} \left[\mathcal{E}\{\|\bar{f}_n - f\|^q_{q,D'}\}\right]^{1/q}$$

be the worst-case risk, measured in the norm $\|\cdot\|_{q,D'}$, of the estimate $\bar{f}_n$ on the class $W_p^l(D,L)$.

Given a set $\mathcal{M}_n$ of values of the "parameter"

$$\mu = (D, L, p, l, q),$$

let us call the quantity

$$\pi(n) = \sup_{\mu=(D,L,p,l,q)\in\mathcal{M}_n} \frac{\mathcal{R}_{q,D',q}(\bar{f}_n; W_p^l(D,L))}{\mathcal{R}^*_{q,D',q}(n; W_p^l(D,L))}$$

the *index of nonoptimality of the estimate* $\bar{f}_n$ *under uncertainty* $\mathcal{M}_n$.

A "good" spatial adaptive estimate should have "small" (ideally – $O(1)$) nonoptimality as $n \to \infty$ under "large" uncertainties $\mathcal{M}_n$. Of course, this is not a formal definition – what are the above "small" and "large"? In our opinion, a reasonable specification here could be as follows. Let us fix once for ever the "structure parameters" l, p, q; for the sake of definiteness, let us even set $p = q$. Let also $pl > 1$ (this is the standard assumption of well-posedness of the problem of nonparametric estimation). With these specifications, the "size" of $\mathcal{M}_n$ is completely characterized by the freedom in L and D:

$$\mathcal{M}_n = \{(D, L, p, l, p) \mid (D, L) \in \mathcal{D}_n\} \tag{54}$$

for certain $\mathcal{D}_n$. Now we may ask whether there exists an estimate with "small" nonoptimality, say, $\pi(n) = O(1)$, with respect to the aforementioned uncertainty. The answer, of course, depends on the "bound" $\mathcal{D}_n$. Note that there exists a "widest reasonable" bound of this type which can be defined as follows. It is known that the optimal risk possesses the standard representation

$$\mathcal{R}^*_{p,D',p}(n; W^l_p(D,L)) = O(1)L^{1/(2l+1)}|D|^{2l/(p(2l+1))}n^{-l/(2l+1)} \tag{55}$$

with $O(1)$ depending on p and l only, provided that

$$(D,L) \in \mathcal{D}^*_n = \{|D|^{1/p}n^l \geq L \geq D^{1/p-l-1/2}n^{-1/2}\}; \tag{56}$$

this latter bound corresponds to the case when the "optimal" estimate based on local polynomial approximation of the signal is related to the window of the size $\geq n^{-1}$ and $\leq |D|$, so that the estimate indeed is optimal in order. Now we may formulate our initial question "whether optimal in order spatial adaptive estimation is possible" as whether there exists an estimator with nonoptimality $\pi(\cdot) = O(1)$ under uncertainty given by the "widest reasonable" bound (56).

As we shall see in a while, the answer to this question is negative – with uncertainty given by (56), the nonoptimality index of any estimation method grows logarithmically with n. Similar phenomenon occurs even for the "smaller uncertainties"; for our goals it is convenient to consider uncertainties

$$(D,L) \in \mathcal{D}^+_n \equiv \{|D| \geq n^{-\gamma},\ \ n^{l-1/p} \geq L \geq |D|^{1/p-l-1/2}\left(\frac{\ln n}{n}\right)^{1/2}\}; \tag{57}$$

here $\gamma \in (0,1)$ is fixed. Note that, for small positive γ, one has $\mathcal{D}^+_n \subset \mathcal{D}^*_n$.

Theorem 5 *For any fixed l, p, γ there exists constant $C > 0$ and N such that the nonoptimality index under uncertainty $\mathcal{M}^+_n$ given by (54) - (57) for any estimation method is $\geq C(\ln n)^{l/(2l+1)}$, provided that $n \geq N$.*

Remark 1 It is easily seen that for the spatial adaptive estimate $\hat{f} = \hat{f}_n$ developed in Section 3 (equipped with $\kappa = c_{m,p}\sqrt{\ln n}$) the nonoptimality index under uncertainty (54) - (57) with $l \leq m$ is $O(1)(\ln n)^{l/(2l+1)}$, so that the estimate, in a sense, possesses the best possible spatial adaptive abilities. To see this, it suffices to note that,

- first, the upper bound on the risk $\mathcal{R}_{p,D',2}(\bar{f}; W^l_p(D,L))$ given by Corollary 2 in the case of (57) reduces to

$$\mathcal{R}_{p,D',2}(\hat{f}, W^l_p(D,L)) \leq O(1)|D|^{2l/(p(2l+1))}L^{1/(2l+1)}\left(\frac{\ln n}{n}\right)^{l/(2l+1)},$$

with $O(1)$ depending on m, p, γ only;

- second, the same estimate is valid also for the risk $\mathcal{R}_{p,D',p}$, provided that $\kappa = c_{m,p}\sqrt{\ln n}$, with properly chosen $c_{m,p}$ (the latter fact is immediately seen from the proofs).

Proof of Theorem 5. The proof is quite similar to that one used by Lepskii [5] to demonstrate that it is impossible to get optimal in order adaptive to smoothness estimator of the value of a signal at a given point. In the below proof C_i denote positive constants depending on γ, l, p only; for any segment D, D', as above, denotes the twice smaller concentric segment.

Let us fix n and $\bar{f} \in F_n$, and let

$$\zeta_n = \frac{n}{\ln n}, \quad \mu_0(n) = (D = D_0 \equiv [0,1], L \equiv L_0 = \zeta_n^{-1/2}, p, l, p);$$

note that $\mu_0(n) \in \mathcal{M}_n^+$ is the "simplest" (with the largest possible D and the smallest, for this D, possible L) value of the "parameter" μ in $\mathcal{M}_n^+$. The signal $f_0 \equiv 0$ belongs to the class $\mathcal{F}_0 \equiv W_p^l(D_0, L_0)$; let $r_0(n)$ be the risk of the estimate $\bar{f}_n$ at this signal:

$$r_0(n) = \left[\mathcal{E}_0\{\|\bar{f}\|_{p,D_0'}^p\}\right]^{1/p},$$

where $\mathcal{E}_0$ is the expectation over the distribution of the observations related to $f \equiv 0$. For $n \geq C_1$ we clearly can find a segment $D_1 \subset D_0'$ of the length $2n^{-\gamma}$ such that

$$\bar{r}_0(n) \equiv \left[\mathcal{E}_0\{\|\bar{f}\|_{p,D_1}^p\}\right]^{1/p} \leq C_2 n^{-\gamma/p} r_0(n).$$

Now set

$$\mu_1(n) = (D = D_1, L = L_1 \equiv |D_1|^{1/p-l-1/2}\zeta_n^{-1/2} = C_3 n^{-\gamma(1/p-l-1/2)}\zeta_n^{-1/2}, p, l, p),$$

so that $\mu_1(n) \in \mathcal{M}_n^+$; this is the "simplest" (with the smallest possible value of L) parameter in $\mathcal{M}_n^+$ with $D = D_1$. Let $\mathcal{F}_1 = W_p^l(D_1, L_1)$, and let

$$r_1(n) = \sup_{f \in \mathcal{F}_1} \left[\mathcal{E}\{\|\bar{f} - f\|_{p,D_1'}^p\}\right]^{1/p}$$

be the uniform risk of the estimate $\bar{f}$ on the class $\mathcal{F}_1$. Let also

$$\rho_0(n) = L_0^{1/(2l+1)}|D_0|^{2l/(p(2l+1))}n^{-l/(2l+1)} = \zeta_n^{-1/(2(2l+1))}n^{-l/(2l+1)}, \tag{58}$$

$$\rho_1(n) = L_1^{1/(2l+1)}|D_1|^{2l/(p(2l+1))}n^{-l/(2l+1)} = C_4 n^{\gamma/2-\gamma/p}n^{-l/(2l+1)}\zeta_n^{-1/(2(2l+1))}. \tag{59}$$

As we know from (55), the quantities $\rho_i(n)$, within constant factor, coincide with the minimax risks associated with the classes $\mathcal{F}_i$, $i = 0, 1$.

Now let

$$\theta = \max\{\frac{r_0(n)}{\rho_0(n)}, \frac{r_1(n)}{\rho_1(n)}\};$$

from the above remarks, the nonoptimality index $\pi(n)$ of the estimator $\bar{f}$ under uncertainty $\mathcal{M}_n^+$ satisfies the inequality

$$\pi(n) \geq C_5\theta. \tag{60}$$

Now let

$$\delta = 8\theta\rho_1(n)n^{\gamma/p}, \tag{61}$$

and let f_1 be the signal which equals to δ on D_1, is between 0 and δ outside D_1 and vanishes outside a little bit larger than D_1 segment D_2. We clearly have $f_1 \in \mathcal{F}_1$ and

$$\|f_1\|_{p,D_1'} = \delta n^{-\gamma/p} = 8\theta\rho_1(n).$$

Now let us define a routine $\mathcal{P}$ which distinguishes between two hypotheses on the distribution of observations: H_0 and H_1; H_i is that the observations relate to the signal $f = f_i$, $i = 0, 1$. The routine is as follows: given observations, we compute $\bar{f}$ and look to which of two signals – f_0 or f_1 – it is closer, in the distance $\|\cdot\|_{p,D_1'}$. Let us evaluate the probabilities p_i to reject the hypothesis H_i in the case when it is valid, $i = 0, 1$.

If H_1 is valid, then the routine will accept H_0 only in the case when the error of estimating f_1 is at least $\frac{1}{2}\|f_1\|_{p,D_1'} = 4\theta\rho_1(n) \geq 4r_1(n)$ (the concluding inequality follows from the definition of θ); since r_1 is $\geq$ the expected error of $\bar{f}$ at f_1 (this is the origin of r_1), we conclude that $p_1 \leq 1/4$.

Similarly, if H_0 is valid, then the routine will accept H_1 only in the case when the error of estimating of f_0, *measured in the norm* $\|\cdot\|_{p,D_1'}$, is $\geq 4\theta\rho_1(n)$. On the other hand, we know that the expected error of estimating f_0 (measured in the norm $\|\cdot\|_{p,D_1}$) is $\leq \bar{r}_0(n)$ – this was the origin of D_1. Consequently,

$$p_0 \leq [4\theta\rho_1(n)]^{-1}\bar{r}_0(n) \leq [4\theta\rho_1(n)]^{-1}C_2 n^{-\gamma/p} r_0(n) \leq C_2 n^{-\gamma/p}\rho_0(n)/\rho_1(n)$$

(we have used the fact that $\theta^{-1}r_0(n) \leq \rho_0(n)$ by the origin of θ). Substituting (58), (59), we get $p_0 \leq C_2 n^{-\gamma/2}$.

Since $p_1 \leq 1/4$ and $p_0 \leq C_2 n^{-\gamma/p}$, the Kullback distance between the distributions of the observations related to f_0 and f_1 is, for $n \geq C_6$, at least $C_7 \ln n$. On the other hand, this distance clearly is $\leq C_8\delta^2 n^{1-\gamma}$, provided that D_2 is close to D_1. Thus, we come to the inequality

$$n^{1-\gamma}\delta^2 \geq C_9 \ln n, \;\; n \geq C_{10};$$

substituting (59) and (61), we come to

$$\theta \geq C_{11}(\ln n)^{l/(2l+1)};$$

this inequality, combined with (60), completes the proof. ■

We see that it "optimal in order" (with $\pi(n) = O(1)$) spatial adaptation under "wide" uncertainties $\mathcal{M}_n$ is impossible. An interesting question would be whether such adaptation is possible under "more narrow" uncertainties of the type

$$|D| \geq n^{-\alpha}, \;\; n^{-\beta} \leq L \leq n^{\gamma}$$

with positive (possibly small) α, β, γ.

7 Implementation and numerical results

This section is devoted to the implementation of the adaptive estimate and the related numerical results. The numerical results are obtained for estimating $f(x)$.

7.1 Implementation

Let us start with some useful observations. To construct $\hat{f}(x_0)$, we consider all Least Square window estimates of the chosen order *m associated with different windows centered at x_0*, and select among them according to certain rule (which is the essence of the matter). It is, however, clear that we could apply the same rule to the Least Square estimates associated with the windows for which x_0 is the right end point, thus coming to another, "forward" estimate, let it be called $\hat{f}_l$; similarly, we could apply the rule to the estimates associated with the windows for which x_0 is the left end point, thus getting a new "backward" estimate $\hat{f}_r$. It is clear from the above proofs that the estimates $\hat{f}_r$ and $\hat{f}_l$ admit the same characterizations of quality, given by Theorems 1 and 2, as the "symmetric" estimate $\hat{f}$. It follows that *any convex combination of the estimates $\hat{f}, \hat{f}_r, \hat{f}_l$*, with the coefficients being arbitrary measurable functions of x and the observations, also satisfies the indicated theorems. With this in mind, we can try to combine our three estimates to improve the practical behavior of the estimator; a pleasant point is that when choosing this combination, we may follow any kind of heuristics without risk of violating the theoretical guarantees of the quality of the estimate given by Theorems 1 and 2.

To make clear the heuristics we actually use, let us speak about the zero order estimates ($m = 0$); for the sake of simplicity, let us speak only about restoring the signal at a point $x^i = i/n$ of the grid, $0 \le i \le n$. According to our construction,

$$\hat{f}(x^i) = S_{i,s(i)}, \quad S_{i,k} = \frac{1}{2k+1} \sum_{j=-k}^{k} y(x^{i+j}),$$

where

$$s(i) = \max \{ s \le \min[i, n-i] \mid \text{ the segments } D_k = [S_{i,k} - \frac{2\kappa}{\sqrt{2k+1}}, S_{i,k} + \frac{2\kappa}{\sqrt{2k+1}}],\ k = 0, ..., s, \text{ have a point in common} \}.$$

In what follows we denote by

$$\rho(i) = (2s(i) + 1)^{-1}$$

the index of the resulting adaptive window at the point x^i. From motivation given in Section 2, this quantity, or, more exactly, κ times this quantity, is "very likely" to be of order of the error of our estimate at x^i.

Similarly,

$$\hat{f}_r(x^i) = S^r_{i,r(i)}, \quad S^r_{i,k} = \frac{1}{k+1} \sum_{j=0}^{k} y(x^{i+j}),$$

$$r(i) = \max \{ r \le n - i \mid \text{ the segments } D^r_j = [S^r_{i,k} - \frac{2\kappa}{\sqrt{k+1}}, S^r_{i,k} + \frac{2\kappa}{\sqrt{k+1}}],\ k = 0, ..., r, \text{ have a point in common} \},$$

$$\hat{f}_l(x^i) = S^l_{i,l(i)}, \quad S^l_{i,k} = \frac{1}{k+1} \sum_{j=0}^{k} y(x^{i-j}),$$

$$l(i) = \max \{ l \le i \mid \text{ the segments } D^l_j = [S^l_{i,k} - \frac{2\kappa}{\sqrt{k+1}}, S^l_{i,k} + \frac{2\kappa}{\sqrt{k+1}}],\ k = 0, ..., l, \text{ have a point in common} \}.$$

The estimates $\hat{f}_r(x^i)$, $\hat{f}_l(x^i)$ also can be equipped with indices, which now are

$$\rho_r(i) = (r(i) + 1)^{-1/2}, \quad \rho_l(i) = (l(i) + 1)^{-1/2},$$

respectively. With our (informal) interpretation of the indices as the measures of quality of the corresponding estimates, it is reasonable to combine the estimates $\hat{f}(x^i)$, $\hat{f}_r(x^i)$, $\hat{f}_l(x^i)$ with the weights inverse proportional to the indices (or certain power of the indices). To make this point clear, look what happens in the following three basic situations: (s): f is "almost constant" in a large symmetric neighborhood of the point x^i; (l): f is "almost constant" to the left of x^i and "varies significantly" to the right of the point; (r): the case symmetric to (l). In case (s) it is preferable to use the symmetric estimate $\hat{f}$, and at the same time its index is very likely to be the smallest of the three indices in question; in case (l) we would prefer to use the "forward" estimate $\hat{f}_l$, and its index is very likely to be the smallest, and similarly in the case (r). Motivated by this simple reasoning, the zero-order estimate we actually used in our experiments was

$$f^*(x^i) = \frac{\rho^{-2}(i)\hat{f}(x^i) + \rho_r^{-2}(i)\hat{f}_r(x^i) + \rho_l^{-2}(i)\hat{f}_l(x^i)}{\rho^{-2}(i) + \rho_r^{-2}(i) + \rho_l^{-2}(i)}. \tag{62}$$

Similar approach was used in the first-order estimate ($m = 1$); the corresponding "symmetric" component and its index clearly are the same as for the zero-order estimate, only the "forward" estimate $\hat{f}_l$ and the "backward" one $\hat{f}_r$ (and their indices) vary in the evident manner.

Note that the arithmetic cost of computing the estimates along all points from the grid is very moderate; with evident preprocessing, it is at most $O(n^2)$.

A difficult, from the practical viewpoint, issue is how to choose the order m of the estimate. Theoretically, there is no problem at all, since the estimate of an order m admits the same, up to depending on m constant factors, accuracy bounds as the estimate of order $m' < m$ on signals of "low" smoothness (with $l \leq m' + 1$) and dominates the estimate of order m' on signals of "high" smoothness ($m' < l \leq m + 1$), so that theoretically a reasonable policy would be to use high-order estimates (with m slowly going to ∞ as n grows). In practice, anyhow, the constant factors are, basically, as crucial as orders, and the choice of the order is not that simple; in fact, when using high order estimates on samples of moderate size, we "oversmooth" the result and do not reproduce properly the "high frequency" components of the signal.

In our opinion, the following decision is appropriate. Let us compute *all* (symmetric, backward and forward) estimates of *all* orders not exceeding a given m, and let us combine all these estimates with the weights inverse proportional to certain power of the indices of the estimates. Note that this approach is *not* good from the theoretical viewpoint, since when mixing low and high order estimates, we cannot ensure quality of the result better than that one for the low order estimates (so that the resulting quality should not be "nearly optimal" for highly smooth signals). In spite of this theoretical drawback, the idea turns out to be rather attractive from the practical viewpoint.

7.2 Numerical results

As the test signals, we used the functions *Blocks*, *Bumps*, *HeaviSine* and *Doppler* given in [1, 2] (see formulas therein[1]). In all experiments, $n = 2048$. In the first series of the experiments (Figures 1.1 - 1.4), same as in [1, 2], the signals f given by the corresponding formulae were, before adding the noise, scaled to have

$$(\frac{1}{n+1}\sum_{i=0}^{n} f^2(x^i))^{1/2} \equiv \chi = 7,$$

[1] in [2], from where we took the description of the test functions, there is a slight discrepancy between the formulae and the figures for *Blocks* and *Bumps*. For *Blocks*, we followed the formulae; for *Bumps* the weights w_i were taken twice smaller

so that the signal-to-noise ratio in this series was 7 (recall that our noise is $\mathcal{N}(0,1)$). In the second series (Figures 2.1 - 2.4), the signal-to-noise ratio was reduced to 3. The below figures present the scaled signal, the observations (signal + noise) and three estimates of the signal, given, respectively, by the adaptive zero-order (0), first-order (1) and "combined" zero-first order (0-1) (see the end of the previous subsection) estimates. When combining the "adaptive components" (symmetric, backward and forward of all orders involved) of the estimate, the weights, same as in (62), were inverse proportional to the squared indices of the components. In all the experiments 2κ was the 10^{-5} quantile of the standard Gaussian distribution ($\kappa = 2.2$).

8 Concluding remarks

To conclude this paper, let us note that the aforementioned results can be extended at least in the following two directions:

- Estimating multivariate regression functions (cf. [8]);
- Estimating functions satisfying linear differential inequalities. Here, as above, the goal is to estimate a signal $f : [0,1] \to \mathbf{R}$ given observations (1), but the "typical" signals which we would like to restore near-optimally satisfy differential inequalities of the type

  $$|Q(\frac{d}{dx})f| \le L, \tag{63}$$

  $Q(z) = z^l + \sum_{i=0}^{l-1} q_i z^i$ being an *unknown*, its own for every signal in question, polynomial of degree not exceeding a given m. Here we are interested in something like Theorem 2, with assumption **A** replaced with

  $\mathbf{A}^*$. There exist $L > 0$, l, $m \ge l \ge 1$, a segment $\Delta \subset [0,1]$ *and a polynomial* Q, $\deg Q = l$, such that (63) is satisfied on Δ.

  Note that assumption **A** with $p = \infty$ is a very particular case of $\mathbf{A}^*$ corresponding to *fixed* and very specific polynomial $Q(z) = z^l$. It turns out that the outlined extension does exist; we plan to investigate these issues in a separate paper.

References

[1] D. Donoho and I. Johnstone. Adapting to unknown smoothness via wavelet shrinkage. Technical report, Department of Statistics, Stanford University, 1992.

[2] D. Donoho and I. Johnstone. Ideal spatial adaptation via wavelet shrinkage. Technical report, Department of Statistics, Stanford University, 1992.

[3] D. Donoho, I. Johnstone, G. Kerkyacharian, and D. Picard. Wavelet shrinkage: Asymptopia? Technical report, Department of Statistics, Stanford University, 1993.

[4] A. Juditsky. Wavelet estimators: Adapting to unknown smoothness. Technical report, IRISA, Rennes, 1994.

[5] O. Lepskii. On a problem of adaptive estimation in gaussian white noise. *Theory of Probability and Its Applications*, 35(3):454–466, 1990.

[6] O. Lepskii. Asymptotically minimax adaptive estimation I: Upper bounds. Optimally adaptive estimates. *Theory of Probability and Its Applications*, 36(4):682–697, 1991.

[7] O. Lepskii, E. Mammen, and V. Spokoiny. Optimal spatial adaptation to inhomogeneous smoothness: an approach based on kernel estimates with variable bandwidth selectors. Technical report, 1994.

[8] A. Nemirovskii. On nonparametric estimation of smooth regression functions. *Sov. J. Comput. Syst. Sci.*, 23(6):1–11, 1985.

[9] M. Pinsker and S. Efroimovich. Learning algorithm for nonparametric filtering. *Automation and Remote Control*, 45(11):1434–1440, 1984.

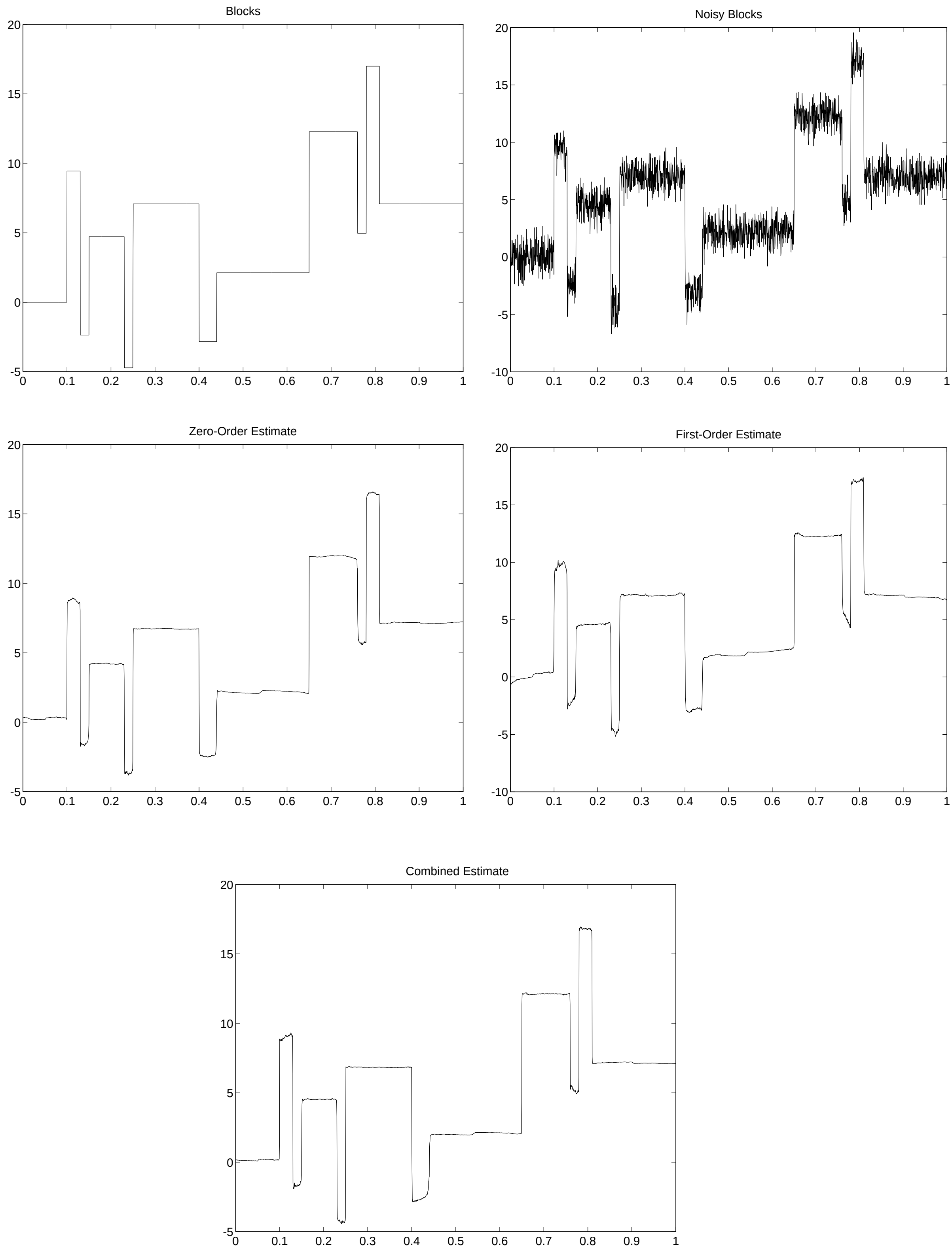
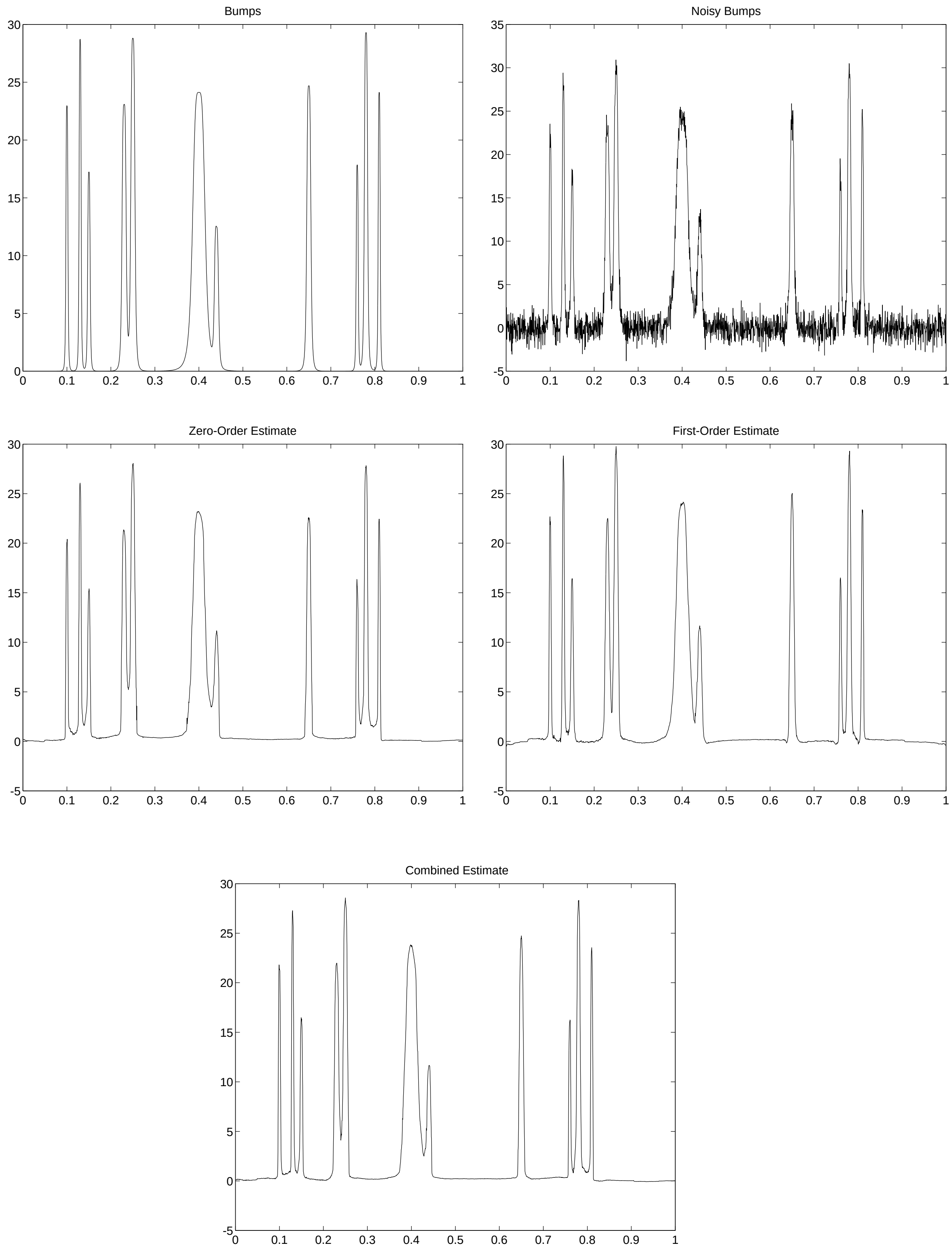

Blocks
Noisy Blocks
Zero-Order Estimate
First-Order Estimate
Combined Estimate

Figure 1.1



Bumps
Noisy Bumps
Zero-Order Estimate
First-Order Estimate
Combined Estimate

Figure 1.2

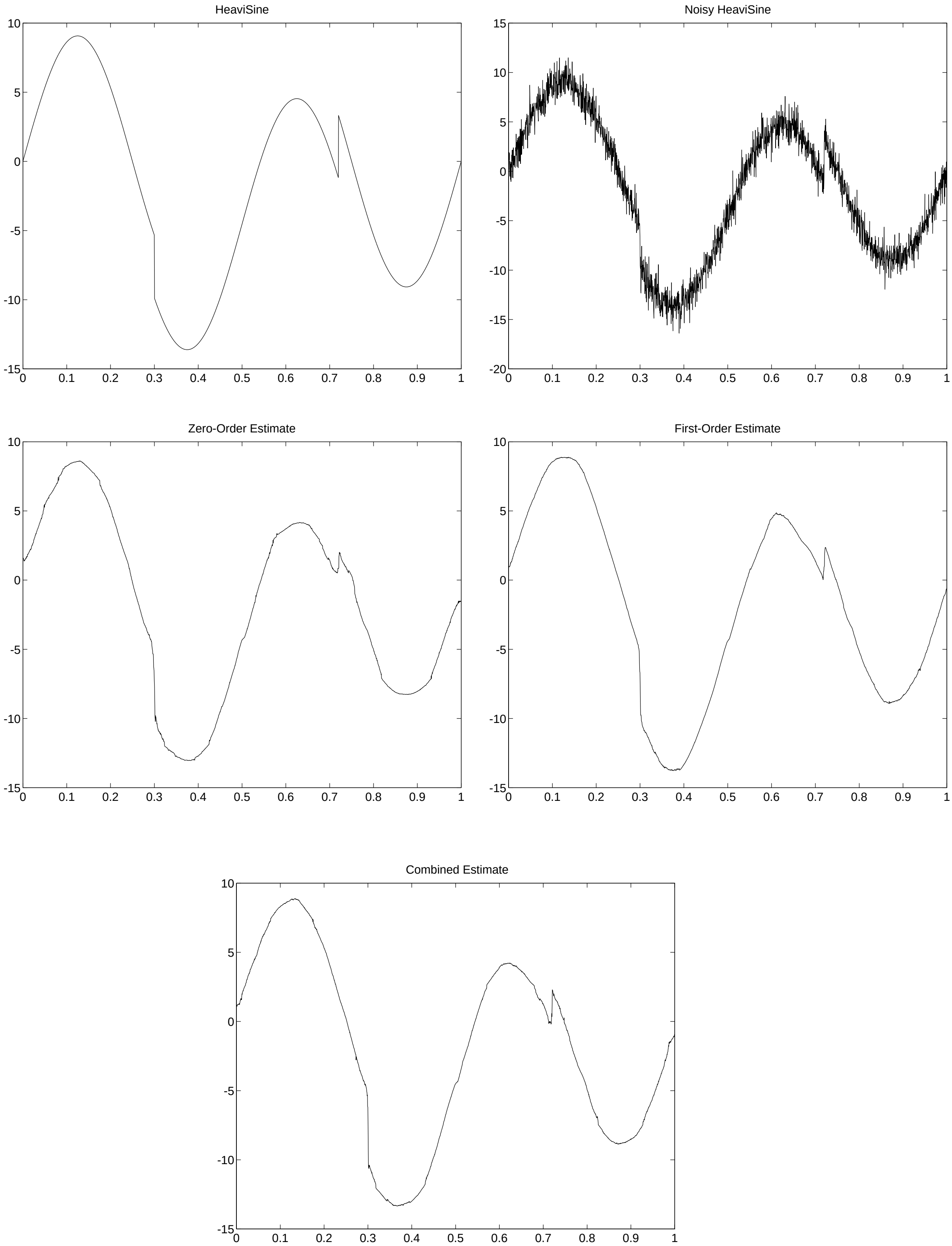

HeaviSine
Noisy HeaviSine
Zero-Order Estimate
First-Order Estimate
Combined Estimate
10
5
0
-5
-10
-15
-20
15
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1

Figure 1.3

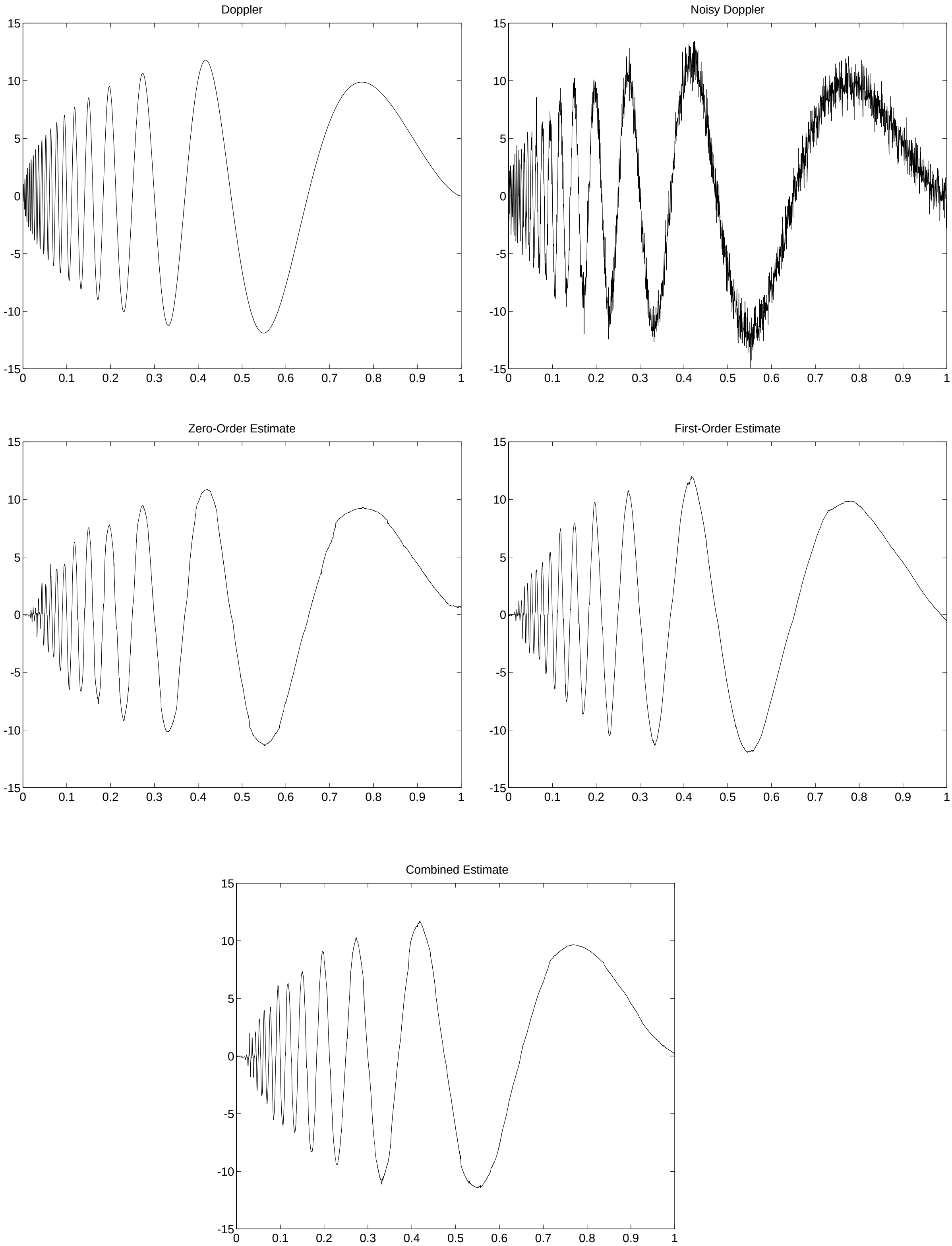

Doppler
Noisy Doppler
Zero-Order Estimate
First-Order Estimate
Combined Estimate
15
10
5
0
-5
-10
-15
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

Figure 1.4

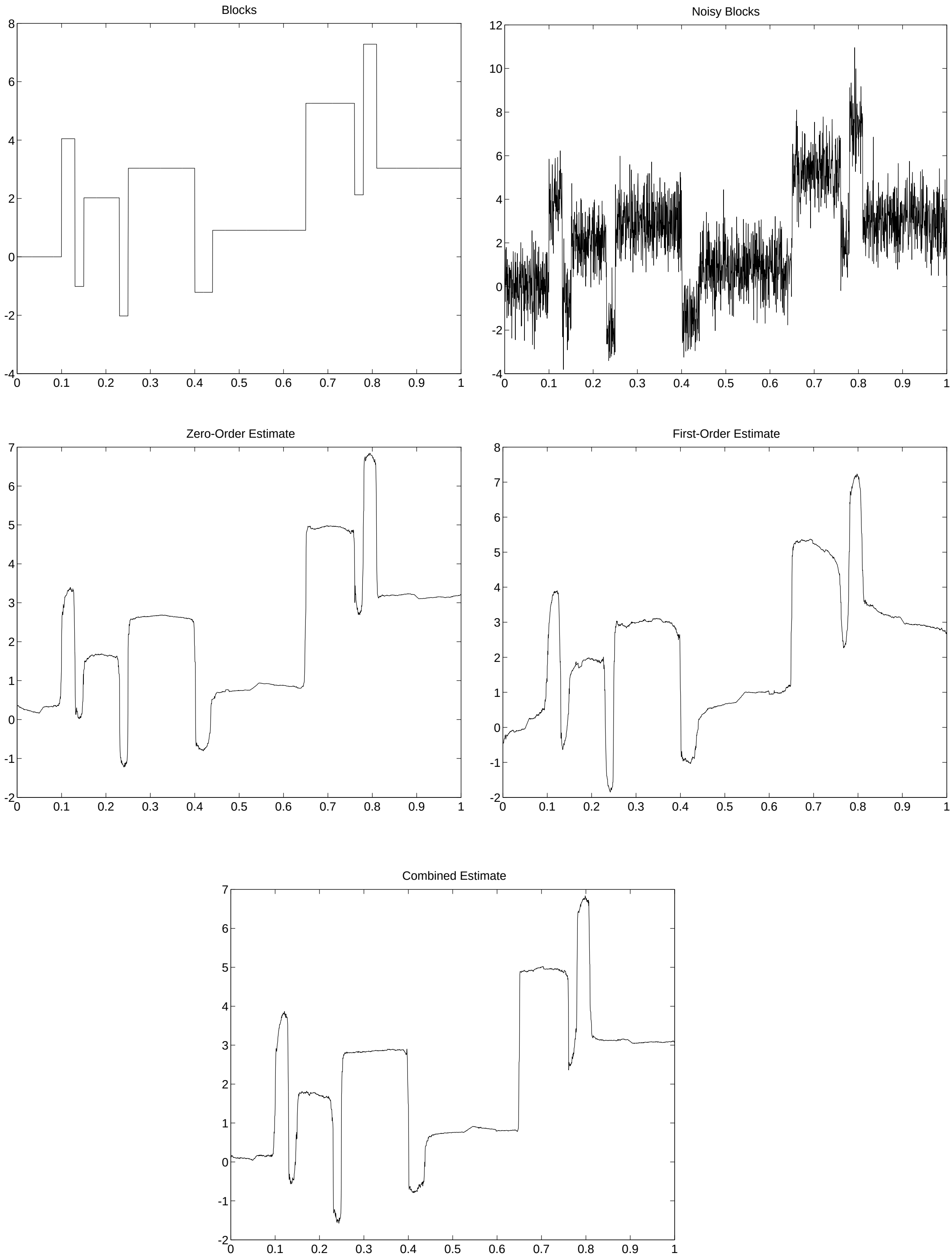

Blocks
8
6
4
2
0
-2
-4
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
Noisy Blocks
12
10
8
6
4
2
0
-2
-4
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
Zero-Order Estimate
7
6
5
4
3
2
1
0
-1
-2
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
First-Order Estimate
8
7
6
5
4
3
2
1
0
-1
-2
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
Combined Estimate
7
6
5
4
3
2
1
0
-1
-2
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1

Figure 2.1

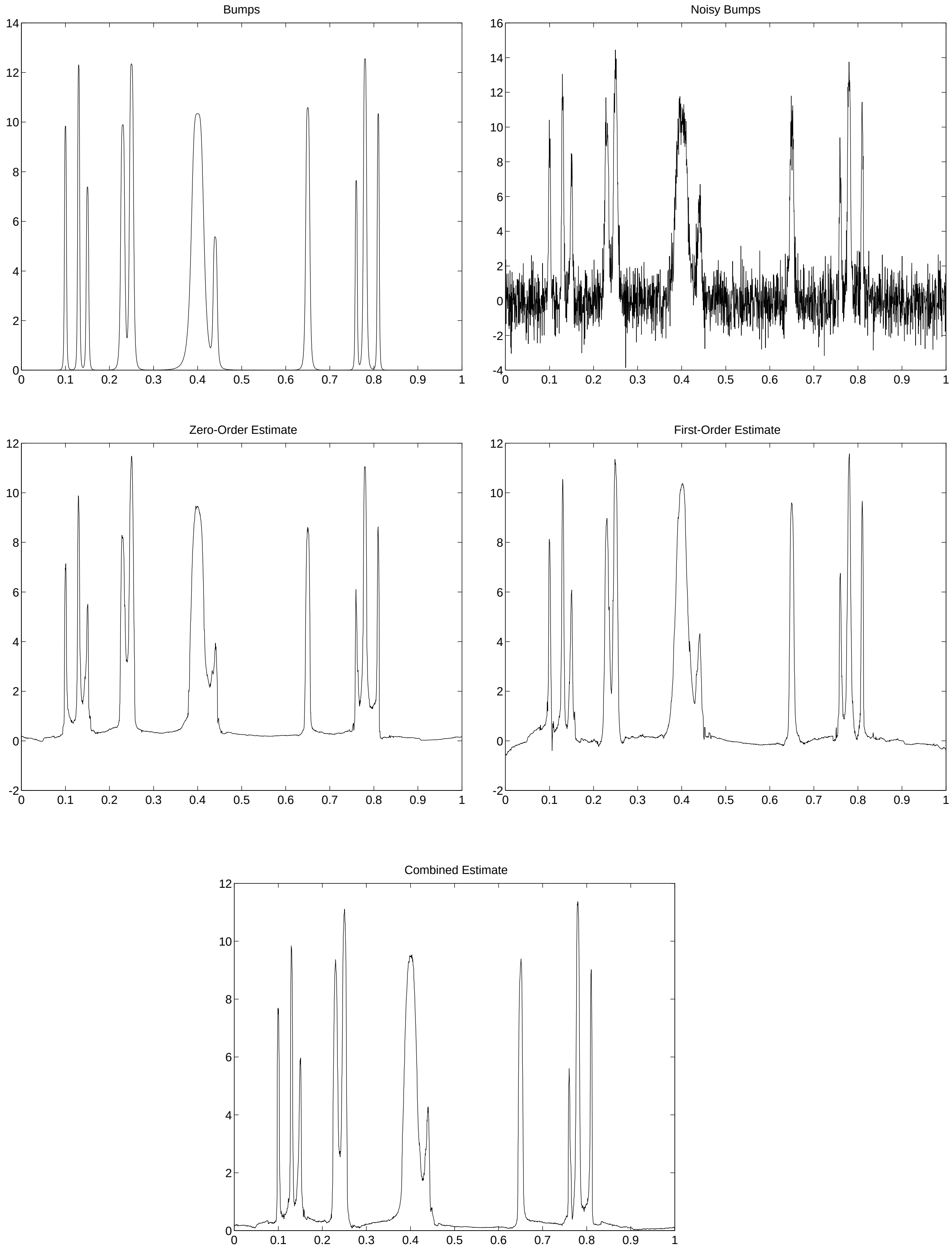

Bumps
Noisy Bumps
Zero-Order Estimate
First-Order Estimate
Combined Estimate

Figure 2.2

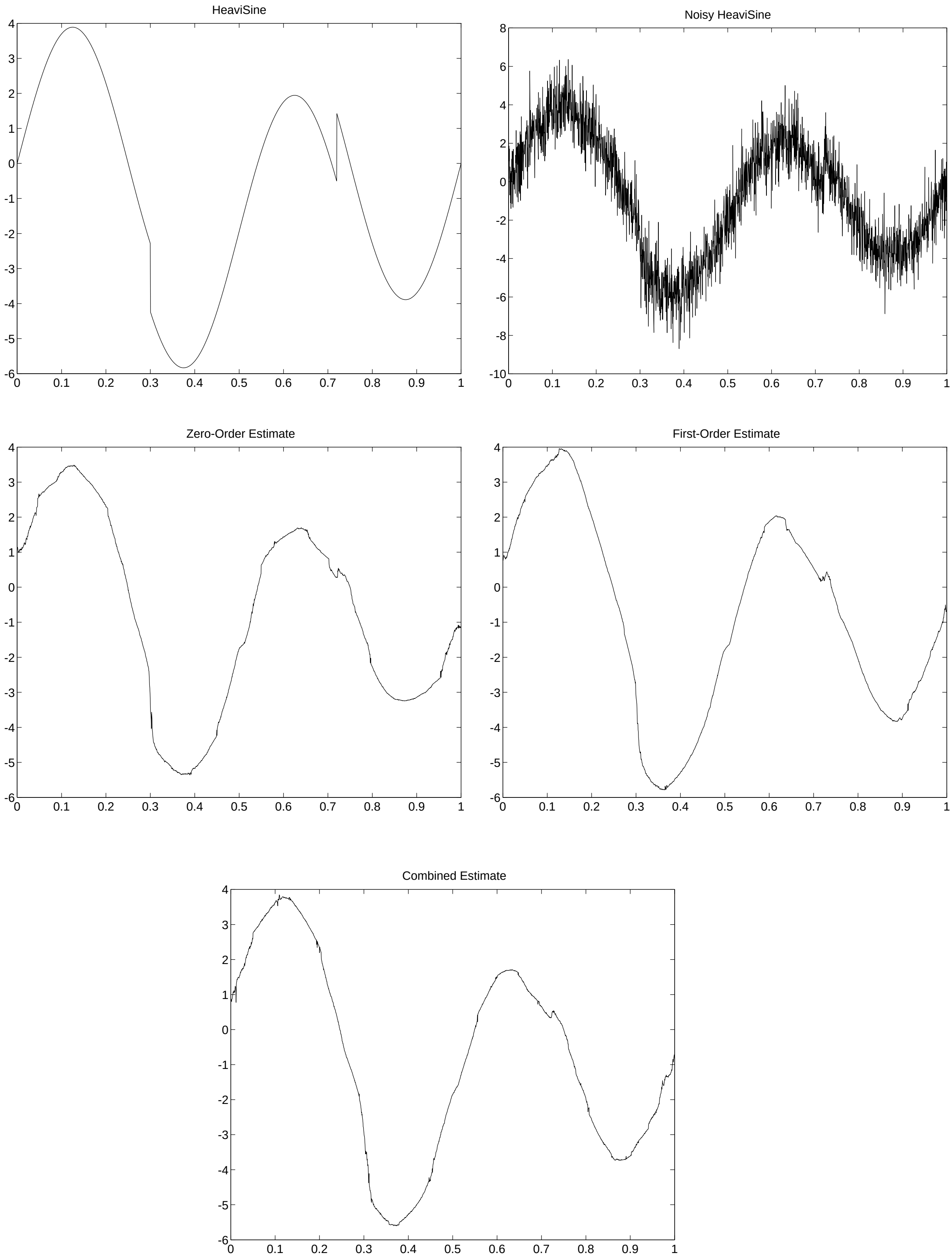

HeaviSine
Noisy HeaviSine
Zero-Order Estimate
First-Order Estimate
Combined Estimate

Figure 2.3

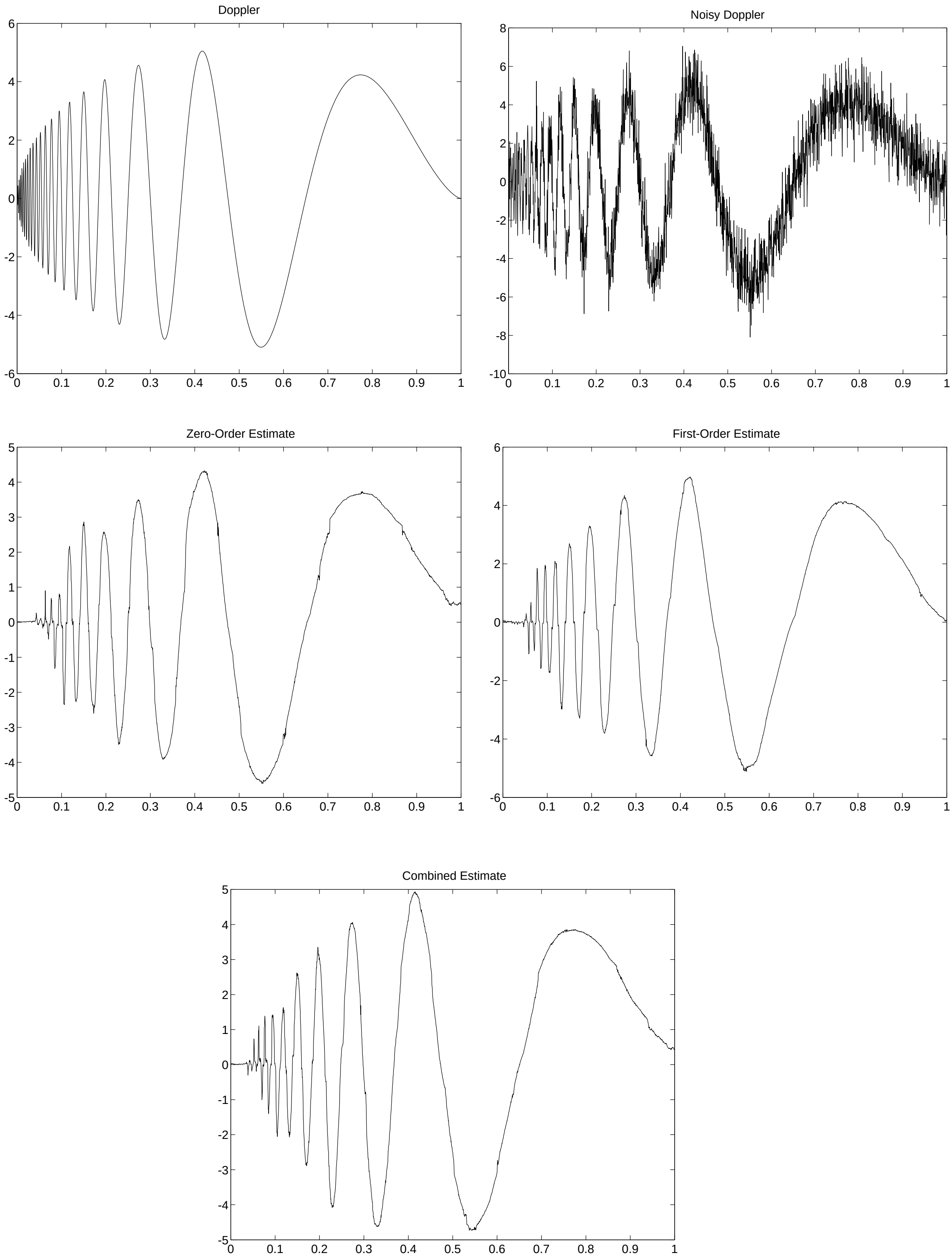

Doppler
Noisy Doppler
Zero-Order Estimate
First-Order Estimate
Combined Estimate

Figure 2.4